



Year: 2017

Targeted proteomics for multiplexed verification of markers of colorectal tumorigenesis

Uzozie, Anuli Christiana ; Selevsek, Nathalie ; Wahlander, Asa ; Nanni, Paolo ; Grossmann, Jonas ; Weber, Achim ; Buffoli, Federico ; Marra, Giancarlo

Abstract: Targeted proteomic methods can accelerate the verification of multiple tumor marker candidates in large series of patient samples. We utilized the targeted approach known as selected / multiple reaction monitoring (SRM/MRM) to verify potential protein markers of colorectal adenoma identified by our group in previous transcriptomic and quantitative shotgun proteomic studies of a large cohort of precancerous colorectal lesions. We developed SRM assays to reproducibly detect and quantify 25 (62.5%) of the 40 selected proteins in an independent series of precancerous and cancerous tissue samples (19 adenoma/normal mucosa pairs; 17 adenocarcinoma/normal mucosa pairs). Twenty-three proteins were significantly upregulated (n=17) or downregulated (n=6) in adenomas and/or adenocarcinomas, as compared with normal mucosa (linear fold changes ± 1.3 , adjusted P value <0.05). Most changes were observed in both tumor types (upregulation of ANP32A, ANXA3, SORD, LDHA, LCN2, NCL, S100A11, SERPINB5, CDV3, OLFM4, and REG4; downregulation of ARF6 and PGM5), and a five-protein biomarker signature distinguished neoplastic tissue from normal mucosa with a maximum area under the receiver operating curve greater than 0.83. Other changes were specific for adenomas (PPA1 and PPA2 upregulation; KCTD12 downregulation) or adenocarcinoma (ANP32B, G6PD, RCN1, and SET upregulation; downregulated AKR1B1, APEX1, and PPA1). Some changes significantly correlated with a few patient- or tumor-related phenotypes. Twenty-two (96%) of the 23 proteins have a potential to be released from the tumors into the bloodstream, and their detectability in plasma has been previously reported. The proteins identified in this study expand the pool of investment-worthy biomarker candidates that can be used to develop a standardized pre-colonoscopy blood test for the early detection of colorectal tumors.

DOI: <https://doi.org/10.1074/mcp.M116.062273>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-130650>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Uzozie, Anuli Christiana; Selevsek, Nathalie; Wahlander, Asa; Nanni, Paolo; Grossmann, Jonas; Weber, Achim; Buffoli, Federico; Marra, Giancarlo (2017). Targeted proteomics for multiplexed verification of markers of colorectal tumorigenesis. *Molecular Cellular Proteomics*, 16(3):407-427.
DOI: <https://doi.org/10.1074/mcp.M116.062273>

Targeted Proteomics for Multiplexed Verification of Markers of Colorectal Tumorigenesis^{*}

Anuli Christiana Uzozie[‡], Nathalie Selevsek[§], Asa Wahlander[§], Paolo Nanni[§], Jonas Grossmann[§], Achim Weber[¶], Federico Buffoli^{||}, and Giancarlo Marrazzi^{**}

Targeted proteomic methods can accelerate the verification of multiple tumor marker candidates in large series of patient samples. We utilized the targeted approach known as selected/multiple reaction monitoring (S/MRM) to verify potential protein markers of colorectal adenoma identified by our group in previous transcriptomic and quantitative shotgun proteomic studies of a large cohort of precancerous colorectal lesions. We developed SRM assays to reproducibly detect and quantify 25 (62.5%) of the 40 selected proteins in an independent series of precancerous and cancerous tissue samples (19 adenoma/normal mucosa pairs; 17 adenocarcinoma/normal mucosa pairs). Twenty-three proteins were significantly up-regulated ($n = 17$) or downregulated ($n = 6$) in adenomas and/or adenocarcinomas, as compared with normal mucosa (linear fold changes $\geq \pm 1.3$, adjusted p value < 0.05). Most changes were observed in both tumor types (up-regulation of ANP32A, ANXA3, SORD, LDHA, LCN2, NCL, S100A11, SERPINB5, CDV3, OLFM4, and REG4; downregulation of ARF6 and PGM5), and a five-protein biomarker signature distinguished neoplastic tissue from normal mucosa with a maximum area under the receiver operating curve greater than 0.83. Other changes were specific for adenomas (PPA1 and PPA2 up-regulation; KCTD12 downregulation) or adenocarcinoma (ANP32B, G6PD, RCN1, and SET up-regulation; downregulated AKR1B1, APEX1, and PPA1). Some changes significantly correlated with a few patient- or tumor-related phenotypes. Twenty-two (96%) of the 23 proteins have a potential to be released from the tumors into the bloodstream, and their detectability in plasma

has been previously reported. The proteins identified in this study expand the pool of biomarker candidates that can be used to develop a standardized precolonoscopy blood test for the early detection of colorectal tumors. *Molecular & Cellular Proteomics* 16: 10.1074/mcp.M116.062273, 407–427, 2017.

Colorectal malignancies are a major driver of cancer-related morbidity and mortality in adults (1, 2). Most of these cancers originate as benign, adenomatous lesions of the colorectal mucosa (3–5). Although they are noninvasive, colorectal adenomas already harbor genetic and epigenetic alterations that disrupt the homeostatic equilibrium between proliferation, differentiation, and apoptosis (particularly those involving components of the Wnt-APC- β catenin pathway). With time, however, some adenomas acquire additional mutations that result in their transformation. Colorectal adenomas and early-stage adenocarcinomas can be eradicated with ease, and the prognosis in these cases is generally good (2, 6, 7).

Early detection is thus the best way to reduce the incidence of colorectal cancer (CRC)¹. Major progress on this front has been hampered, however, by the limitations of available screening techniques (8–10). Colonoscopy offers high diagnostic accuracy and allows prompt removal of suspicious early-stage lesions, but endoscopic screening is invasive, costly, and time-consuming. Fecal assays have none of the latter drawbacks, but their specificity is low, and they are reportedly less effective in the detection of adenomas than of adenocarcinomas (10–12).

One of the most promising solutions for filling this diagnostic gap involves the use of targeted proteomic methods to identify and quantify tumor-associated proteins in body fluids.

From the [‡]Institute of Molecular Cancer Research, University of Zurich, Switzerland; [§]Functional Genomics Center Zurich, University/ETH Zurich, Zurich, Switzerland; [¶]Institute of Surgical Pathology, University of Zurich, Switzerland; ^{||}Gastroenterology and Endoscopy Unit, Hospital of Cremona, Italy

Received, July 8, 2016, and in revised form, January 4, 2017

Published, MCP Papers in Press, January 4, 2017, DOI 10.1074/mcp.M116.062273

Author contributions: A.C.U., P.N., F.B., and G.M. designed research; A.C.U., N.S., A. Wahlander, and A. Weber performed research; N.S., A. Wahlander, and P.N. contributed new reagents or analytic tools; A.C.U. and J.G. analyzed data; A.C.U. and G.M. wrote the paper; A. Weber histologic diagnostic revision; F.B. clinical sample collection.

¹ The abbreviations used are: CRC, colorectal cancer; ANOVA, analysis of variance; AUC, area under the curve; CE, collision energy; CEA, carcinoembryonic antigen; Cscore, combined discriminate score; ELISA, enzyme-linked immunosorbent assay; FDR, false discovery rate; ICC, intraclass correlation coefficient; iRT, indexed retention time; iTRAQ, isobaric tags for relative and absolute quantification; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MS, mass spectrometry; ROC, receiver operating characteristics; SRM, selected reaction monitoring; TFA, trifluoroacetic acid.

Major ongoing advances in proteomics technology are providing researchers with methods that are more sensitive and reproducible than data-dependent techniques (13–16). Selected reaction monitoring (SRM)—also known as multiple reaction monitoring (MRM)—is a sensitive targeted proteomic technique for the reproducible quantification of specific proteins within a complex sample background (14, 17, 18). Importantly, in a single SRM experiment, numerous precursor/fragment ion pairs (*i.e.* transitions), and therefore several proteins can be monitored across a relatively large number of samples. The resulting chromatographic trace provides information on the retention time and signal intensity per transition, which can be used to determine the relative or absolute abundance of several target peptides. If results are normalized to a defined amount of a stable isotope-labeled internal peptide, endogenous protein abundance can also be compared in various samples and/or under diverse test conditions (17, 19, 20).

Proteomic studies have already pinpointed several potential biomarker candidates for the detection of CRC. However, most studies stop short of verifying the candidates in a systematic and reproducible manner, an essential step for translating biomarker discoveries to clinical application (6, 8, 9, 21). Verification, however, is an arduous process, which is made even more difficult by the lack of appropriate tools for identifying and reproducibly quantifying levels of candidate proteins across multiple samples (6, 21). Enzyme-linked immunosorbent assays (ELISAs) are still the preferred methods for pre-clinical verification of candidate biomarkers, but the antibodies used for these assays are characterized by high specificity requirements and elevated production costs, which render them less than ideal for testing large numbers of tumor biomarkers (19, 22). Targeted proteomic techniques have been proposed as standard complementary methods for the clinical verification and validation of candidate disease biomarkers, and these methods are being used in an increasing number of clinical proteomic-based cancer studies (19, 23–33).

In a comprehensive shotgun proteomics study of human colorectal adenomas and patient-matched samples of normal colon mucosa, we recently identified several putative diagnostic markers representing tumor stages along the adenoma-adenocarcinoma pathway (34). In the present study, we developed targeted SRM assays (27) and used them to profile the expression of these candidate protein markers in a large, independent group of colorectal tumors, precancerous and invasive. The results revealed potential markers that significantly distinguished CRC tumors from healthy tissues, showed substantial correlation with clinical parameters and are apparently detectable in human plasma.

EXPERIMENTAL PROCEDURES

Human Tissue Samples—Tissue samples were used with local ethics committee approval and in conformity with the principles set forth in the Declaration of Helsinki. Human colorectal tissues were

collected prospectively during colonoscopies performed at the Istituto Ospitalieri di Cremona, Italy. All donors provided written consent to sample collection, testing, and data publication. Samples were identified by numerical codes to protect donors' rights to confidentiality. The series comprised 19 adenomas and 17 adenocarcinomas, each with a matched sample of normal mucosa from the colon segment bearing the tumor (>2 cm from the lesion). The characteristics of the paired normal and tumor tissue samples are summarized in Table I. Immediately after collection, samples were frozen in liquid nitrogen and stored at -80°C .

Protein Extraction from Tissues—Rapidly weighed frozen tissue samples (weight range: 10 to 70 mg) were placed in solution (180 μl) containing 100 mM triethylammonium bicarbonate (Sigma, St. Louis, MO, USA), 1X Complete Ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Mixture (Roche, Mannheim, Germany), 1 M urea, 5 mM β -glycerophosphate disodium salt hydrate, 1 mM sodium orthovanadate, and 5 mM sodium fluoride (Sigma) and homogenized on ice in a Wheaton glass borosilicate grinder. The homogenate was transferred to a 1.5-ml Eppendorf tube, the grinder was washed with 60 μl of lysis buffer, and the combined volumes were sonicated with a Bioruptor (Diagenode, Denville, NJ) (high power, five 10"/10" ON/OFF cycles) and centrifuged (16,000 $\times g$ for 5 min at 4°C). The supernatant containing the tissue protein extract was collected and stored at -80°C .

Protein Digestion—For each sample, 50 μg of protein in denaturing buffer (initial reaction volume = 62.5 μl in 4 M Urea) were reduced with 10 mM dithiothreitol (1.25 μl of a 500-mM stock solution) at 35°C for 45 min and alkylated with 50 mM iodoacetamide (6.4 μl of a 500-mM stock solution) at room temperature for 1 h in the dark. Dithiothreitol (50 mM: 7 μl of a 500-mM solution) was added to terminate the alkylation reaction, and the samples were incubated at room temperature for 10 min. Afterward, 5 μl of 0.1 $\mu\text{g}/\mu\text{l}$ sequencing-grade endoproteinase Lys-C (Roche) was added to each sample and incubated at room temperature for 3 h. After Lys-C digestion, the samples were diluted with 50 mM ammonium bicarbonate to a final urea concentration of 1.4 M, and digested with 2.5 μl of 0.4 $\mu\text{g}/\mu\text{l}$ sequencing-grade modified trypsin (Promega, Madison, WI) at 37°C for 15 h. To stop the digestion, we added acetonitrile and trifluoroacetic acid (TFA) to a total concentration of 3% acetonitrile and 0.1% TFA, and pH 2–3. The pH of acidified samples was verified with pH-indicator strips (McolorpHast™, Merck, KGaA, Darmstadt, Germany). Compared with in-solution trypsin digestion, serial proteolysis using Lys-C and trypsin increases the yield of fully cleaved peptides, thereby improving the accuracy and sensitivity of SRM protein quantification (35). Peptide solutions were desalted on SepPak C18 columns, lyophilized in a vacuum centrifuge, and resolubilized in 50 μl (1:1 v/w) of a solution of 3% acetonitrile, 0.1% formic acid, and indexed Retention Time (iRT) 10 \times Mix (Biognosys AG, Zurich, Switzerland) (36). The iRT mix was prepared in accordance with the manufacturer's protocol.

Stable Isotope-labeled Standard Peptides (Heavy Peptides) for SRM Assay Library—Isotope-labeled heavy peptides corresponding to the proteotypic peptides selected for the study (supplemental Table S1), and containing either a C-terminal (13C(6) 15N(4)) arginine or a (13C(6) 15N(2)) lysine residue were purchased from JPT Peptide Technologies GmbH (Berlin, Germany). A pool containing known quantities of each standard peptide in the sample matrix solution was subjected to nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis carried out on an LTQ Orbitrap Velos (Thermo Fischer Scientific, Bremen, Germany). Mascot output .dat files were imported into Skyline (37) and used to generate the spectral library for SRM transitions.

SRM Assay Development and Time-scheduled SRM Analysis—The SRM assay was developed on a TSQ Vantage Triple Quadrupole

Mass Spectrometer (Thermo Fischer Scientific) using a sample mixture containing defined amounts of each isotope-labeled peptide and data from the spectral library. The spectral library consisted of transitions measured in data-independent mode in the sample mixture. A preliminary SRM-transition list for at least two peptides per protein was created from the spectral library in Skyline. For each peptide, we selected four to eight dominant transitions (precursor and most intense fragment ion pair) consisting of doubly and triply charged y - and b -ion precursors (2+ and 3+ charge states), and fragment ions with charge states of 1+ and 2+. Peptides were analyzed on the TSQ Vantage in the SRM mode, which involves repeated cycling through a list of transitions, with predefined dwell-times on each transition (17). Retention times extracted from the SRM scans were used to calculate iRT values (36) for each peptide in Skyline. SRM scans were searched to confirm the near absence of endogenous transitions in the crude peptide mix. We selected transitions with optimum SRM properties (14, 17) and iRT values for defining time-scheduled SRM methods to monitor our target proteins. In time-scheduled SRM mode, the full cycle time is used to detect and quantify peptides expected to elute within a given retention-time window. This restricts the acquisition of defined transitions to a window around the elution time of the corresponding peptide. We optimized SRM parameters, including collision energy (CE), as previously described (38), dwell time, and crude peptide spike-in amount. Retention time scheduling for all transitions was selected based on Skyline scheduling predictions. A window of ± 2.5 min was used to schedule all transitions. The final transition list consisted of 600 transitions. To facilitate the detection of low-abundance peptides without compromising the dwell time of 0.020 s per transition, the 600 transitions were divided into two groups of 300 (based on protein name) and analyzed with two different time-scheduled SRM acquisition methods (Method 1, Method 2). Complete information on the two groups of transitions is provided in [supplemental Tables S2 and S3](#). Two proteins (GAPDH and HSP90AA1) generally used as “housekeeping” proteins were included in our protein target list to explore their utility as global standards in colorectal tissues, but they were not used to normalize the intensities of target transitions. SRM transitions for these proteins and iRT peptides were monitored in each method.

Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)—Scheduled SRM measurements were performed on the TSQ Vantage instrument equipped with a nanoelectrospray ionization source and coupled to a NanoLC Ultra 2D HPLC system (Eksigent Technologies, Dublin, CA). The instrument was used in the SRM acquisition mode with the following settings: Q1 resolution, 0.7 FWHM Q3 resolution, 0.7 fwhm; cycle time, 2 s. A spray voltage of 1.3 keV was used with an ion transfer tube set heated to a temperature of 270 °C. Argon was used as the collision gas at a nominal pressure of 1.5 mTorr. Chromatographic separation of peptides was performed on a frit column (150 mm \times 75 μ m) made in-house and coupled to a fused silica emitter (100 mm \times 75 μ m) (New Objective Inc., Woburn, MA). The columns were packed with reverse-phase C18 material (AQ, 3 μ m 200 Å, Bischoff GmbH, Leonberg, Germany) and maintained at 50 °C with an automatic heater during all SRM experiments. Peptides were loaded onto the column from a cooled (4 °C) Eksigent autosampler and separated with a linear gradient of acetonitrile/water containing 0.1% formic acid at a flow rate of 300 nL/min. An elution gradient of 5% to 35% acetonitrile over 60 min was used. Four microliters of sample, corresponding to 1.25 μ g peptide mass, was injected. Blank runs were performed between SRM measurements of paired biological samples to avoid sample carryover.

Sample Randomization and Blocking—Our experimental design employed a block-randomized sample strategy (39). As shown in Fig. 1, the order in which the adenoma/normal mucosa and adenocarcinoma/normal mucosa pairs were analyzed was determined randomly,

and balanced allocation in each group was maintained during both the sample preparation and SRM analysis phases. Proteins were extracted from four tissue samples at a time (i.e. 1 adenoma/normal mucosa pair and 1 adenocarcinoma/normal mucosa pair). For the digestion of the protein extracts, we divided the 36 sample pairs into three groups. One sample pair randomly selected from each group was digested twice to control for unwanted variations in peptide intensity at this step of sample preparation. As a result, the total number of samples for scheduled SRM analysis increased from 72 to 78. The order of spectral acquisition on the TSQ Vantage was also randomized. For each SRM method (Fig. 1), we created a sample queue using the random function in Microsoft Excel. Samples were randomized based on protein digestion group, protein extraction pairs, day of sample processing, and SRM data acquisition. Each patient sample was analyzed with both time-scheduled methods, thus producing 156 acquired data files. Sample analysis time was 270 h, including run times for iRT and blank controls. SRM data was statistically analyzed to detect differences between the adenoma and normal mucosa groups and between the adenocarcinoma and normal mucosa groups.

Data Analysis—SRM data were analyzed as outlined in Fig. 1. Peak group identification and peptide scoring were performed using mProphet, a computational tool for statistically validating SRM mass spectrometry data (40), in SpectroDive software (Version 5.5.5478.20997, Biognosys AG). Input files for this analysis consisted of the data files containing the SRM spectra acquired for each sample during the scheduled runs and the transition assay list detailing specified parameters such as Q1, Q3, CE, iRT score and relative fragment intensity ([supplemental Table S4](#)). SpectroDive processes SRM data using the assay list of target and heavy standard peptides and applies mProphet’s extraction and scoring strategy. mProphet employs a “decoy-transition” approach to probabilistically score individual features in SRM peaks and integrates them into a combined discriminant score (Cscore) while controlling for the false discovery rate (FDR) (40). Decoy transitions are transitions for peptide species that are known to be absent in the biological sample; they therefore function as negative controls. We employed synthetic decoys because our workflow involved the use of isotope-labeled internal standards ([supplemental Table S4](#)). Peptides that were detected in at least 6 samples and had an FDR <1% corresponding to a Cscore >10 were further processed for quantification ([supplemental Table S5](#)).

Protein significance was analyzed with the MSstats R-package (MSstats.daily 2.1.6) in R statistical software (version 3.2.3) (41). The input file for this analysis contained values for *Condition*, *Bioreplicate*, and *Run*, as pre-assigned in Skyline according to the experimental design ([supplemental Tables S6 and S7](#)). Data were first processed in MSstats, and all transition intensities were transformed into \log_2 values. Next, normalization based on “equalizeMedians” (41) was performed on all transitions using the transition intensities of the isotope-labeled standard peptides. This equalized the median of standard intensities across runs and applied similar between-run shifts to endogenous peptide intensities in the experiment. Quantification of protein abundance and analysis of differential abundance in patient groups were performed with the linear mixed effect model for SRM workflows utilizing stable isotope-labeled standard peptides (41, 42) to normalize the intensities of respective endogenous peptides. The resulting p values were corrected for multiple comparisons (41). Significant tumor-associated dysregulation was defined as differential expression in neoplastic (adenoma and/or adenocarcinoma) and normal mucosa characterized by an adjusted p value of <0.05 and a linear fold change $\geq \pm 1.3$ (corresponding to a \log_2 fold change of 0.4).

Statistical and Functional Network Analysis—Protein abundance ratios (endogenous/isotope-labeled standard) were visualized using

the TreeView software (43). The pROC package in R was used to generate receiver operating characteristics (ROC) curves and to calculate areas under the curves (AUCs) and corresponding confidence intervals. The 95% confidence interval of the AUC was computed with 2000 stratified bootstrap replicates for each case. Relationships between protein abundance and clinical variables were assessed with linear models (lm) using single stratum analysis of variance (ANOVA) (categorical variables) or regression (numerical variables) analysis, both performed in R. Intraclass correlation coefficients (ICC) (44) were calculated in R statistical software to assess the agreement between protein expression changes measured with SRM and those obtained in our previous study (34) with iTRAQ (isobaric tags for relative and absolute quantification) methods. Deconvoluted protein ratios were calculated to compare SRM- and iTRAQ-based measurements on the same scale. The deconvoluted protein ratios (e.g. normal mucosa versus adenoma or adenoma versus normal mucosa) derived from SRM measurements were calculated as the ratio of endogenous peptide (L) to heavy peptide (H) in each normal mucosa versus adenoma pair (i.e. normal L/normal H:adenoma L/adenoma H). Deconvoluted iTRAQ protein ratios for normal mucosa versus adenoma were calculated as previously described (34).

Functional network analysis to identify potential interaction partners was performed with the GeneMANIA algorithm (genemania.org). The methods used for weighting and for analyzing network associations were automatically selected in GeneMANIA. Bar plots and dot plots were prepared in GraphPad Prism (version 6). Protein association with diseases was examined using the web version of FunDO (45). Briefly, the disease association of each gene in the genome was annotated using the Disease Ontology and peer-reviewed evidence from GeneRIF (46). A condensed version of the Disease Ontology, Disease Ontology Lite, was used for the statistical analysis (47), and the significance of each disease association was evaluated by Fisher's exact test (48).

RESULTS

SRM Assays for Verification of Candidate Colorectal Tumor Markers—We performed an independent validation study to verify findings from our earlier discovery-based transcriptomic (49) and shotgun proteomic (34) studies of a large cohort of precancerous colorectal lesions. Our objective was to create a high-throughput SRM-based workflow for detecting and quantitating the relative abundance of candidate biomarkers in an independent series of colorectal adenomas. We also examined colorectal adenocarcinomas to validate protein abundance changes during tumor progression. The characteristics of the 19 adenomas and 17 adenocarcinomas analyzed are listed in Table I. Protein abundance in each tumor was normalized to that found in patient-matched samples of normal mucosa. The 40 target proteins selected for this study (supplemental Table S1) included: 29 candidate biomarkers of colorectal neoplasia identified in our previous shotgun proteomics study (34); six proteins putatively involved in the same or related pathways as some of these 29 proteins; three candidate biomarkers of colorectal neoplasia identified in a previous transcriptomic study by our group (49); and two proteins widely used as housekeeping proteins (GAPDH and HSP90AA1). Proteotypic peptide sequences for each protein were selected based on the results of our previous study (34), and from other proteomics data sets available in the Peptide

Atlas database (50). Additionally, peptide sequences conformed with empirical selection criteria including the length of the tryptic peptide sequence, hydrophobicity, uniqueness, and absence of missed cleavages (17, 51, 52). When possible, we also avoided the use of sequences with amino acids prone to chemical modification. This selection resulted in a total of 140 proteotypic peptides, with a minimum of two peptides per protein (supplemental Table S1).

Targeted SRM assays for the 40 selected proteins in unfractionated extracts were developed and optimized as illustrated in supplemental Fig. S1 (see also *Experimental Section*). Although sample fractionation improves sensitivity when combined with shotgun-proteomic workflows, it is less useful in SRM workflows (23, 53). The additional sample preparation steps increase the chances of variation in the measurement of protein abundance. The sample throughput level also declines because multiple fractions from each sample have to be measured individually.

Stable isotope-labeled heavy standard peptides with amino acid sequences identical to those of the corresponding 140 endogenous peptides were synthesized and analyzed by LC-MS/MS to validate the peptide sequence. This step was necessary to ensure confident identification of each peptide monitored, to compensate for suppression effects, and to improve measurement precision (27). Spectral libraries were generated in Skyline with data acquired by shotgun analysis of a mixture of pooled aliquots of the 140 crude synthetic peptides.

Of the 140 peptides analyzed, only 74, corresponding to 31 proteins (supplemental Table S8), were detected. Nondetection of the other peptides may have been related to several factors, including suboptimal MS ionization properties and/or matrix effects caused by other components in the unfractionated samples.

Targeted Analyses of Colorectal Tissue Samples by LC-SRM—The optimized SRM assays were used to quantify 31 target proteins in colorectal tissue samples. Tissue peptide digests were assayed in randomized order (Fig. 1). Equal amounts of each standard peptide were introduced into each tissue peptide sample from a pooled mixture of 74 standard peptides. Standard isotope-labeled peptides were used as internal controls during statistical analysis of our SRM data and for the measurement of relative changes in protein abundance across tissue samples. Each tissue peptide digest also contained a known concentration of a mixture of iRT reference peptides (36), which served as internal controls for monitoring instrument and technical variations during time-scheduled SRM measurements. Based on precursor and fragment ion intensities, two to four transitions were selected for each standard (heavy) and endogenous (light) peptide.

Reproducibility of SRM Measurements—The sensitivity and validity of SRM measurements can be affected by pre-analytical variability (introduced by factors such as sample prep-

TABLE I

Characteristics of the neoplastic lesions included in the study. Abbreviations: AD, adenoma; ADC, adenocarcinoma; C, cecum; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid; R, rectum; TA, tubular adenoma; TVA, tubulovillous adenoma; VA, villous adenoma; SSA, sessile serrated adenoma; NA, data not available (as for cancers, staging and grading not available since patient was treated in a different hospital)

Patient no.	Sample type/no. ^a	Age	Sex	Location (Colon segment)	Maximum diameter (mm)	Paris classification ^b	Pit pattern ^c	Microscopic appearance	Highest grade of dysplasia ^d	Cancer stage/grade ^e	Lesions excised at study colonoscopy ^f	Lesions excised during prior colonoscopy ^g
1	AD-1	50	F	C	30	0 – IIa	IIIL	TA	low	-	1	1
2	AD-2	72	F	A	20	Is	IV	TVA	low	-	1	0
3	AD-3	66	F	S	35	Ip	IV	TA	low	-	1	0
4	AD-4	78	M	R	50	IIa + Is	IV + Vi	TVA	low	-	1	0
5	AD-5	58	F	R	20	IIa + IIc	IIIL – Vi	TA	low	-	2	0
6	AD-6	60	M	C	30	Is	IV	TVA	low	-	2	2
7	AD-7	65	M	A	40	IIa + Is	IV	TA	low	-	1	0
8	AD-8	70	F	R	20	Is + IIa	IV	TVA	low	-	1	0
9	AD-9	70	M	A	40	Is + IIa	IV	TA	low	-	1	0
10	AD-10	77	F	R	80	IIa + Is	IV	VA	low	-	1	0
11	AD-11	83	M	S	20	IIa + Is	IIIL	TVA	low	-	3	0
12	AD-12	75	F	S	30	IIa + Is	IIIL – IV	TA	low	-	1	3
13	AD-13	63	M	T	30	0 – Is	IV	TVA	low	-	1	0
14	AD-14	74	F	A	80	Is + IIa	IV	TVA	low	-	1	0
15	AD-15	82	M	D	50	Is + IIa	IV	VA	low	-	1	0
16	AD-16	61	F	R	40	Is + IIa	II	SSA	low	-	1	0
17	AD-17	72	F	C	45	0 – Is	IV	TA	high	-	1	0
18	AD-18	58	M	D	20	0 – Ip	IV	TA	high	-	2	0
19	AD-19	83	F	C	35	Is	IV	TVA	high	-	1	0
20	ADC-1	78	M	R	60	IIa + Is	Vi – IV	cancer	-	NA	1	1
21	ADC-2	58	M	S	40	Is	IV + Vi	cancer	-	NA	2	1
22	ADC-3	67	F	R	40	0 – Is	Vn	cancer	-	NA	1	0
23	ADC-4	37	M	R	60	0 – Is	Vn	cancer	-	NA	1	0
24	ADC-5	63	M	S	80	0 – Is	Vn	cancer	-	NA	1	0
25	ADC-6	82	M	S	65	0 – Is	Vn	cancer	-	T2N0M0/G2	1	0
26	ADC-7	52	M	S	60	0 – Is	Vn	cancer	-	T4N0M0/G2	1	0
27	ADC-8	56	F	A	65	0 – Is	Vn	cancer	-	NA	1	0
28	ADC-9	55	M	R	70	0 – Is	Vn	cancer	-	T3N0M0/G2	3	0
29	ADC-10	87	F	T	70	0 – Is	Vn	cancer	-	T3N0M0/G2	1	0
30	ADC-11	64	F	A	60	0 – Is	Vn	cancer	-	NA	1	0
31	ADC-12	64	M	C	70	Is	NA	cancer	-	T2N0M0/G2	1	0
32	ADC-13	88	F	A	55	0 – Is	Vi	cancer	-	T3N0M0/G2	1	0
33	ADC-14	67	M	T	80	0 – Is	Vn	cancer	-	T3N1M0/G2	1	0
34	ADC-15	64	M	R-S	100	0 – Is	Vn	cancer	-	NA	2	1
35	ADC-16	61	F	T	60	0 – Is	Vn	cancer	-	T3N0M0/G2	1	0
36	ADC-17	53	F	A	50	0 – Is	Vn	cancer	-	T3N0M0/G2	1	0

^a serial number for AD and ADC lesions.

^b Macroscopic appearance classified according to the Paris Endoscopic Classification of Superficial Neoplastic Lesions. *Gastrointest. Endosc.*, 2003; 58, S3–S43.

^c Kudo classification of colon crypt morphology. Kudo S, Rubio CA, Teixeira CR, et al. Pit pattern in colorectal neoplasia: endoscopic magnifying view. *Endoscopy* 2001; 33, 367–7.

^d WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6–9, 1999. IARC.

^e Cancer TNM and grade classification in according to Sobin LH, Wittekind C. TNM classification of malignant tumors. 6th ed. New York, NY: Wiley-Liss, 2002.

^f Total no. of lesions detected and excised during the study colonoscopy, including the single lesion used in the proteomics analysis.

^g Total no. of lesions detected and excised during previous colonoscopies.

aration) and/or analytical variability (resulting from undesirable variation of instrument parameters during SRM data acquisition) (19). Our block-randomized (39) study design (Fig. 1, explained in Experimental Procedures) ensured adequate sample randomization from the processing phase through SRM spectral acquisition. The paired feature (normal and tumor) of colorectal tissue for each patient was maintained in our randomized design. We examined our SRM data for

unwanted bias caused by variability in instrument performance.

As shown in supplemental Fig. S2, a retention time shift of less than 1 min was maintained for each iRT peptide in all samples analyzed by time-scheduled SRM. Although each tissue sample was analyzed with both time-scheduled methods, chromatographic performance remained stable across all 156 SRM measurements.

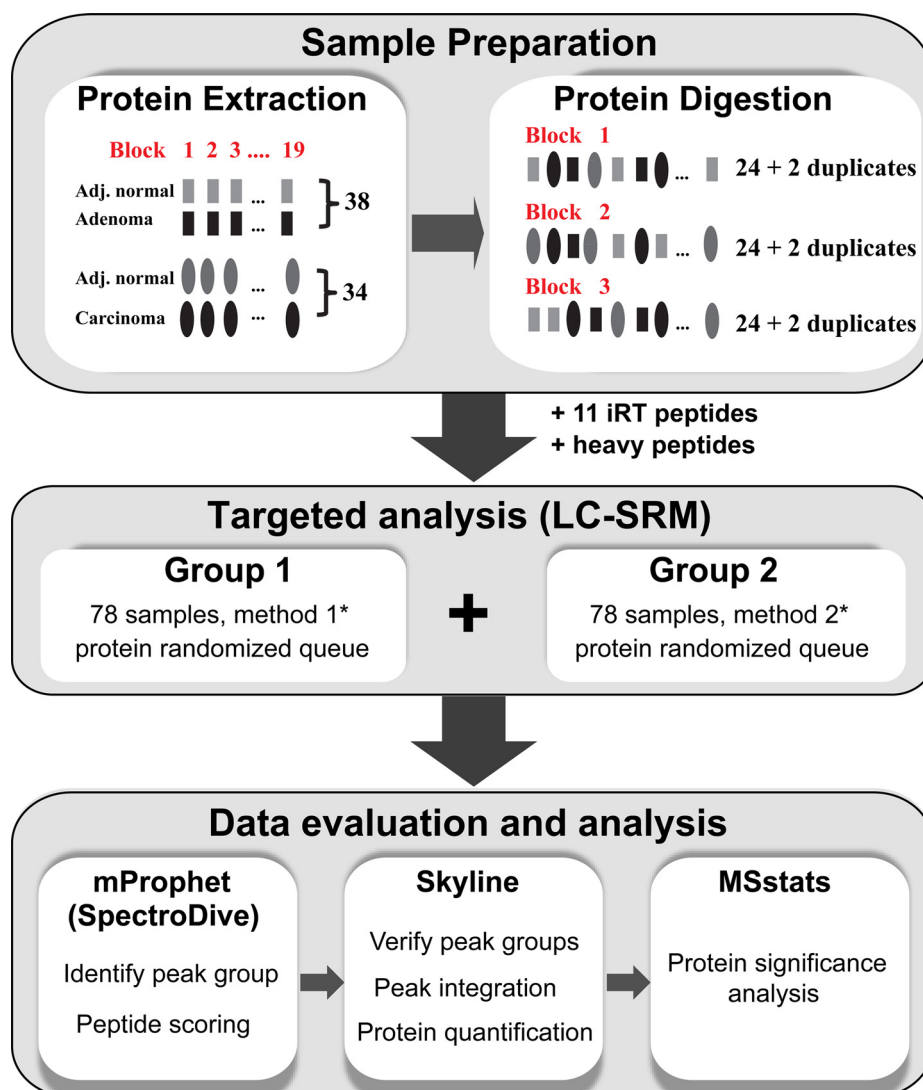


FIG. 1. Block-randomized experimental design for SRM analysis. Extraction of protein from samples was randomized based on the tumor type. Two paired samples (1 consisting of adenoma and adjacent normal mucosa, the other of adenocarcinoma and adjacent normal mucosa) were extracted at the same time. Enzymatic digestion of protein extracts was performed in three batches, each comprising 24 randomly allocated samples. In each batch, one randomly selected matched tissue pair (tumor/normal mucosa) was digested twice (duplicates). Transitions for time-scheduled SRM were divided into two groups, and each sample was analyzed twice according to a randomized queue. More details in *Experimental Section and Results*.

The reproducibility of our SRM spectral peak measurements for each peptide was also assessed by comparing the endogenous (light) to standard (heavy) peptide (L/H) peak ratios for each peptide monitored in duplicate peptide digests from the same protein extract. As shown in Fig. 1, protein extracts from three paired samples of normal and tumor tissues (total: 6 samples) were digested twice, and the resulting 12 digests were analyzed separately on the mass spectrometer, applying a randomized SRM sequence. For each duplicate pair, an R^2 value of >0.9 and an intercept of ~ 1 was observed, indicating that a reproducible peak area was obtained for all peptides measured in each sample duplicate (supplemental Fig. S3). This demon-

strates that the confounding effects of instrument variance on our data and the inferred results were adequately minimized.

Peptide Identification and Quantification of Protein Abundance in Colorectal Tissues—Confident identification of peptides in the 72 tissue peptide digests analyzed was achieved using the mProphet statistical tool (40). A comprehensive list of peptide identification parameters is provided in supplemental Table S5. Peptides were considered identified if peak groups had a Cscore (40) >10 at a controlled FDR (Qvalue) cutoff of $<1\%$. Applying this stringent cutoff (supplemental Table S5), we identified 53 endogenous peptides and achieved a sensitivity of 71.5% in peak group selection. Moreover, we manually

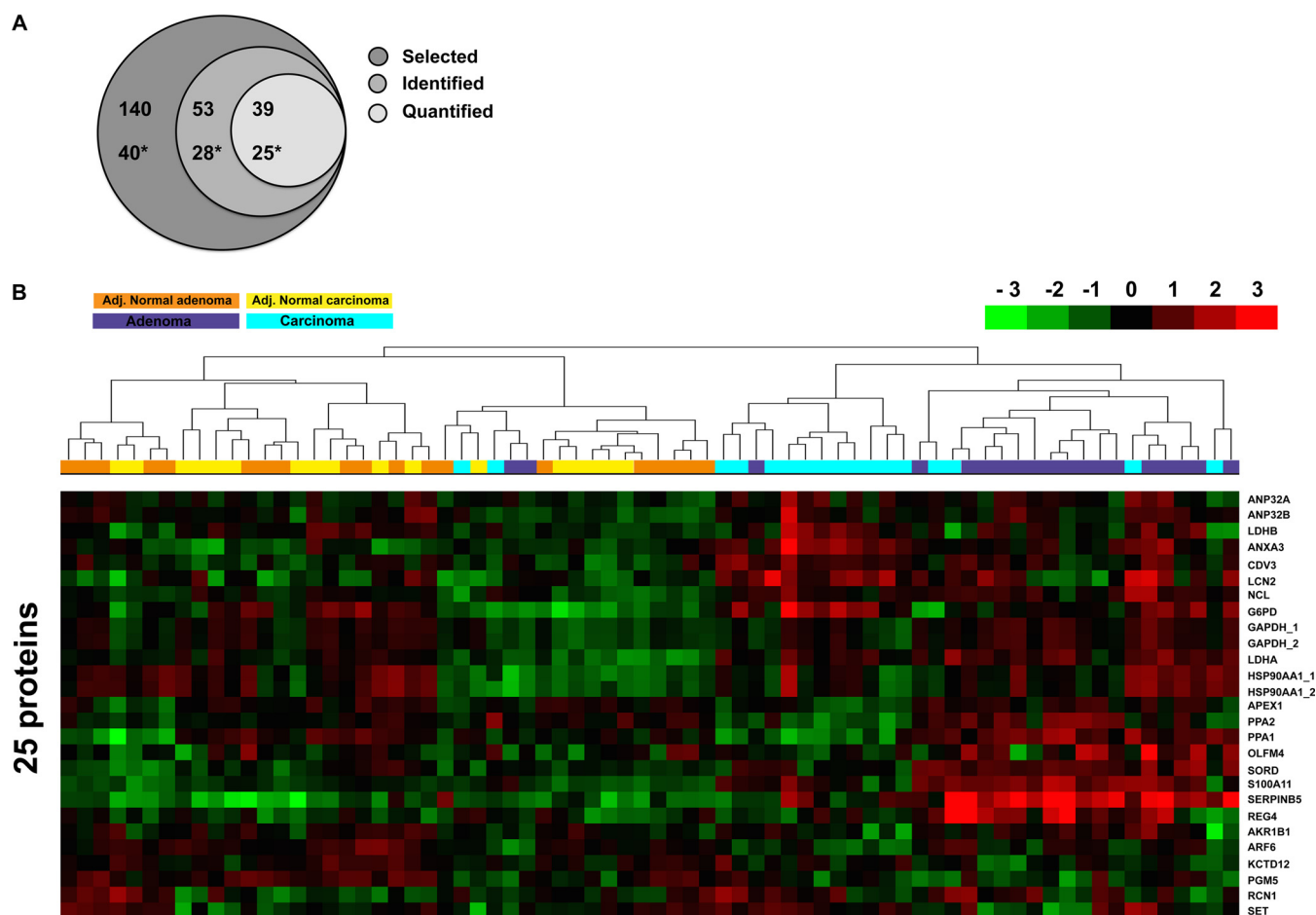


FIG. 2. Targeted quantification of peptide species and proteins. A, Shaded circles represent the number of candidate peptides and proteins selected, identified, and consistently quantified in normal and neoplastic lesions. Protein number is indicated with *. B, The measured data (log transformed light/heavy ratio) for the 25 protein targets quantified. Hierarchical clustering with uncentered correlation and complete linkage was employed to group samples based on measured protein abundance. Tumor samples labeled in *light blue* are adenocarcinomas; those in *dark blue* are adenomas. Adenoma-paired samples of normal mucosa (adjacent normal adenoma) are labeled in *orange*; carcinoma-paired samples of normal mucosa (adjacent normal carcinoma) are in *yellow*.

confirmed the coelution of heavy and light peptides for these 53 peptides.

A peptide peak group was chosen for protein quantification only if the corresponding standard synthetic peptide transitions were observed in peptide digests from all 72 tissue samples analyzed. The 39 peptides that fulfilled this requirement were used to quantify 25 proteins. Overall, the SRM assays developed and refined in the present study to quantify selected proteins in colorectal tissues achieved a success rate of 62.5% (25 out of 40 proteins) (Fig. 2A and supplemental Fig. S4).

SRM Verification of Putative Biomarkers for Colorectal Neoplasia—Our SRM assays successfully detected 25 of the 40 candidate protein biomarkers in unfractionated digests of the tissue proteome. We subsequently confirmed that these assays could be used to reproducibly quantify levels of those 25 proteins across tumor samples and to identify their differential abundance between tumor and normal mucosa samples. All

25 proteins were reliably quantified across samples (Fig. 2B). Statistical analysis to identify significantly dysregulated proteins was performed with MSstats (41, 42), as described in Experimental Procedures. Proteins with an adjusted *p* value of <0.05 and a fold change $\geq \pm 1.3$ were considered significantly altered (Fig. 3). One of the housekeeping proteins, GAPDH, displayed significantly different abundance level in adenomas and normal mucosa samples after the data were normalized with heavy peptides (equalizedMedians option in MSstats, see Experimental Procedures). Therefore, normalizing data to “housekeeping” proteins did not appear to be a reliable method for maximizing quantification of colorectal tumor-related protein variations in this study.

Seventeen proteins showed significant differences in abundance in colorectal *adenoma versus* normal mucosa samples (Fig. 3A, and Table II). ANP32A, ANXA3, SORD, GAPDH, LDHA, LCN2, NCL, S100A11, SERPINB5, CDV3, PPA1, PPA2, OLFM4, and REG4 were up-regulated in adenoma

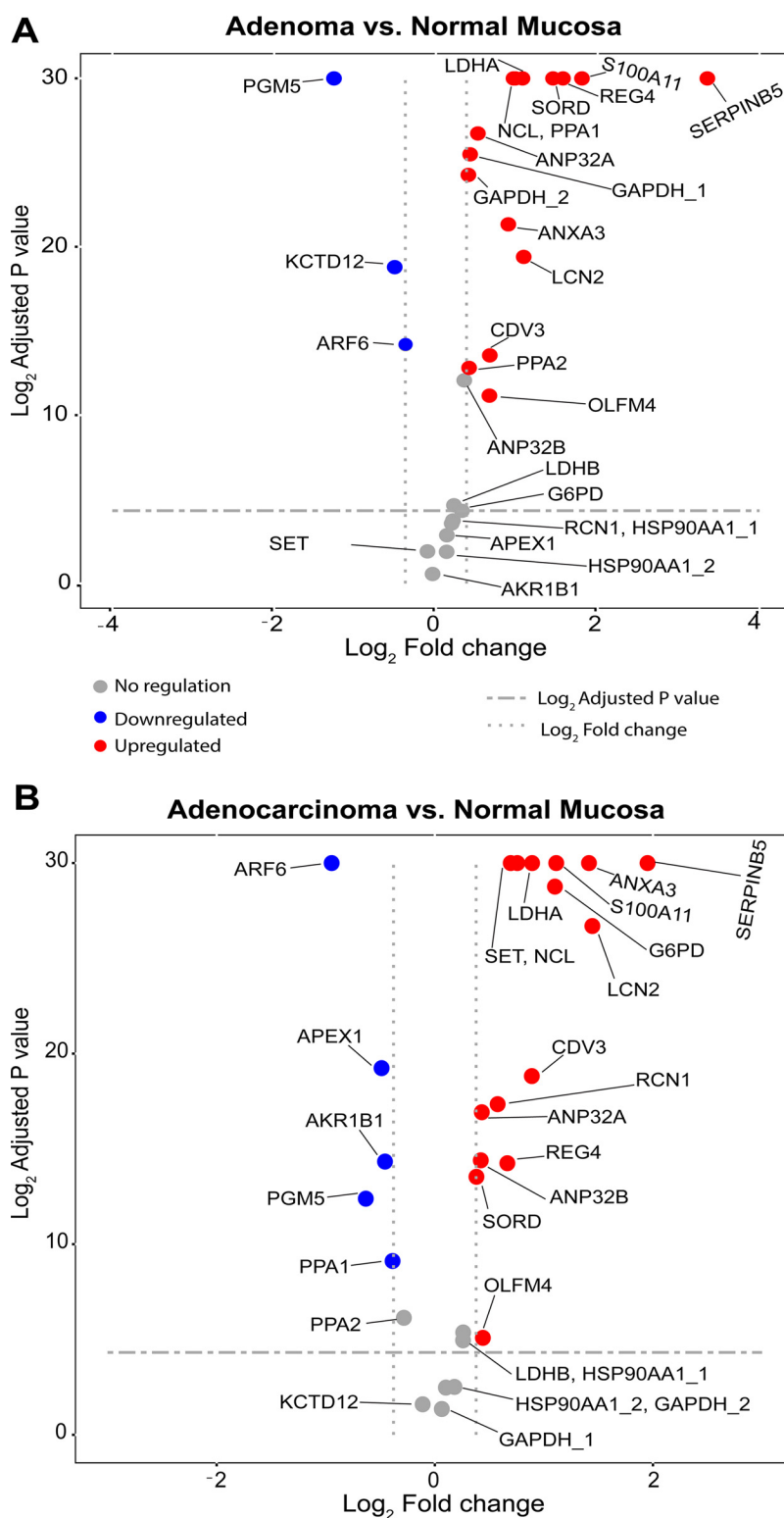


FIG. 3. Quantification of tumor markers in colorectal adenomas (A) and adenocarcinomas (B) (relative to matched normal mucosal samples). Proteins with an adjusted p value < 0.05 and a fold change $\geq \pm 1.3$ were considered significant. Housekeeping proteins monitored in the two time-scheduled methods are denoted with $_1$ and $_2$.

TABLE II
SRM-based quantification of target proteins in colorectal adenomas and adenocarcinomas. Proteins upregulated (↑) or downregulated (↓) in colorectal adenomas and/or adenocarcinomas compared to the normal mucosa. Significantly dysregulated proteins met the cutoff criteria: fold change ≥ 1.3 , adjusted p value < 0.05 . NS: not significant or not detected; UP: upregulated; DOWN: downregulated

Uniprot Accession	Protein name	Gene name	Adenoma vs. Normal mucosa				Adenocarcinoma vs. Normal mucosa				Proteins detected in human plasma/serum				Proteins secreted via classical and nonclassical pathways		
			Log2 Fold change	Adjusted p value	Significant	Trend in iTRAQ data (34)	Log2 Fold change	Adjusted p value	Significant	Human Plasma Proteome (79)	Human Plasma Proteome (80)	Cancer associated proteins in 82 ^c and/or 45 ^d	signalP v4.1 (75)	secretomeP v2.0a (76)	Extra-cellular vesicles (77)		
P15121	ALDR	AKR1B1	0.0	1.03	0.67648	No	-0.5	1.37	0.00005	Yes ↓	Yes	Yes ^{cd}	No	No	Yes		
P39887	AN32A ^a	ANP32A	0.5	1.43	0.00000	Yes ↑	0.4	1.35	0.00001	Yes ↑	Yes	No	No	No	Yes		
Q92688	AN32B ^a	ANP32B	0.4	1.27	0.00023	No	0.4	1.34	0.00005	Yes ↑	Yes	No	No	No	Yes		
P12429	ANXA3	ANXA3	0.9	1.86	0.00000	Yes ↑	1.4	2.66	0.00000	Yes ↑	Yes	No	No	No	Yes		
P27695	APEX1	APEX1	0.1	1.10	0.13701	No	-0.5	1.40	0.00000	Yes ↓	Yes	Yes ^{cd}	No	Yes	Yes		
P62330	ARF6	ARF6	-0.4	1.30	0.00005	Yes ↓	-0.9	1.93	0.00000	Yes ↓	No	Yes ^d	No	Yes	Yes		
Q00796	DHSO	SORD	1.4	2.73	0.00000	Yes ↑	0.4	1.30	0.00008	Yes ↑	Yes	Yes ^c	No	No	Yes		
P04406	G3P ^b	GAPDH	0.4	1.34	0.00000	Yes ↑	0.1	1.05	0.39098	No	Yes	No	No	No	Yes		
P11413	G6PD	G6PD	0.3	1.25	0.05023	No	1.1	2.15	0.00000	Yes ↑	Yes	Yes ^c	No	No	Yes		
P07900	HS90A ^b	HSP90AA1	0.2	1.14	0.08444	No	0.3	1.20	0.03215	No	Yes	Yes ^c	No	No	No		
P00338	LDHA ^a	LDHA	1.1	2.10	0.00000	Yes ↑	0.9	1.86	0.00000	Yes ↑	Yes	Yes ^c	No	No	Yes		
P07195	LDHB ^a	LDHB	0.2	1.17	0.03996	No	0.3	1.20	0.02418	No	Yes	Yes ^d	No	No	Yes		
P80188	NUCL	LCN2	1.1	2.12	0.00000	Yes ↑	1.4	2.72	0.00000	Yes ↑	No	Yes ^d	Yes	Yes ^e	Yes		
P19338	NUCL	NCL	1.0	1.98	0.00000	Yes ↑	0.8	1.69	0.00000	Yes ↑	Yes	Yes ^d	No	No	Yes		
P31949	S10A11	S100A11	1.8	3.49	0.00000	Yes ↑	1.1	2.16	0.00000	Yes ↑	Yes	Yes ^d	No	Yes	Yes		
P36952	SPB5	SERP/INB5	3.4	10.23	0.00000	Yes ↑	2.0	3.87	0.00000	Yes ↑	Yes	No	No	No	Yes		
Q9UKY7	CDV3	CDV3	0.7	1.59	0.00008	Yes ↑	0.9	1.85	0.00000	Yes ↑	Yes	No	No	No	Yes		
Q15181	IPYR ^a	PPA1	1.0	1.94	0.00000	Yes ↑	-0.4	1.31	0.00180	Yes ↓	No	No	No	No	Yes		
Q9H2U2	IPYR2 ^a	PPA2	0.4	1.33	0.00014	Yes ↑	-0.3	1.22	0.01418	No	Yes	Yes ^c	No	No	Yes		
Q96CX2	KCD12	KCTD12	-0.5	1.42	0.00000	Yes ↓	-0.1	1.08	0.32967	No	Yes	No	No	No	Yes		
Q6UX06	OLFM4	OLFM4	0.7	1.58	0.00044	Yes ↓	0.4	1.36	0.02941	Yes ↑	No	Yes ^d	Yes	Yes ^a	Yes		
Q15124	PGM5	PGM5	-1.3	2.40	0.00000	Yes ↓	-0.6	1.55	0.00019	Yes ↓	Yes	No	No	No ^e	Yes		
Q15293	RCN1	RCN1	0.2	1.16	0.07462	No	0.6	1.49	0.00001	Yes ↑	Yes	No	Yes	Yes ^e	Yes		
Q9BYZ8	REG4	REG4	1.6	2.97	0.00000	Yes ↑	0.7	1.59	0.00005	Yes ↑	No	No	Yes	Yes ^e	Yes		
Q01105	SET	SET	-0.1	1.08	0.26687	No	0.70	1.62	0.00000	Yes ↑	Yes	Yes ^d	No	No	Yes		

^a Protein isozymes and isoforms.

^b Data from the first method (method 1) used for “housekeeping” proteins measured in both scheduled SRM methods (methods 1 and 2). Data sets obtained with the two methods were identical; details in Experimental Procedures and Results sections.

^c Cancer associated proteins listed in Polanski and Anderson (82).

^d Cancer associated proteins identified using functional ontology annotations (reference 45).

^e Protein has a signal peptide. SignalP cut-off: D-score ≥ 0.45 , SecretomeP cut-off: NN-score ≥ 0.6 . Extra-cellular vesicles (reference 77): curated proteins in Vesiclepedia detected in microvesicles and exosomes from colorectal cancer cells.

versus normal mucosa samples. By contrast, expression levels of ARF6, KCTD12, and PGM5 were significantly reduced in adenoma samples.

Marked adenocarcinoma-related alterations were observed in the expression of 20 proteins: ANP32A, ANP32B, ANXA3, SORD, G6PD, LDHA, LCN2, NCL, S100A11, SERPINB5, CDV3, OLFM4, RCN1, REG4, and SET were significantly up-regulated, and AKR1B1, ARF6, APEX1, PPA1, and PGM5 were significantly downregulated relative to normal mucosa samples (Fig. 3B, and Table II).

The diagnostic potential of the five most markedly dysregulated proteins (selected on the basis of p values) in each tumor type was assessed using ROC analysis on protein intensity ratios (L/H). Fig. 4 depicts the trade-off between sensitivity and specificity for five adenoma-associated proteins (S100A11, SORD, SERPINB5, NCL, and PGM5) with AUC values ≥ 0.87 . A similar analysis of adenocarcinoma samples yielded AUC values > 0.80 for SERPINB5, ARF6, ANXA3, CDV3, and NCL, indicating a fairly good capacity for distinguishing cancer lesions from normal mucosa (Fig. 5).

In summary, most of the proteins whose expression levels were significantly altered in tumor samples showed directionally similar forms of dysregulation (as compared with normal mucosa) in both adenomas and adenocarcinomas (Table II, supplemental Fig. S5): 11 of these proteins (ANP32A, ANXA3, SORD, LDHA, LCN2, NCL, S100A11, SERPINB5, CDV3, OLFM4, and REG4) were significantly up-regulated, whereas expression levels of ARF6, and PGM5 were decreased in both tumor types. However, some proteins showed significant expression changes in only one of the two tumor types: up-regulation of PPA1 and PPA2 and downregulation of KCTD12 were seen exclusively in adenomas, whereas up-regulation of ANP32B, G6PD, RCN1, and SET and down-regulation of AKR1B1, APEX1, and PPA1 were specific to adenocarcinomas. In $\sim 80\%$ of both types of tumors the direction of dysregulation (increased or decreased protein levels) ($r = 0.77$, p value = 0.001) was the same in both tumor groups. PPA1 was an exception, displaying higher-than-normal abundance in adenomas and lower-than-normal abundance in carcinomas (supplemental Figs. S5 and S6). In a combined ROC analysis using adenoma and adenocarcinomas as a single classifier (i.e. neoplastic tissue), a panel of five proteins (S100A11, SERPINB5, NCL, SORD, and ANXA3) discriminated between normal mucosa and neoplastic tissue with sensitivity ranging from 70% to 100% at 80% specificity (Fig. 6).

We investigated the correlation (p values and Pearson's r score) between the abundance of these five proteins and different clinical features of the tumors using linear models for regression analysis or ANOVA (see Experimental Procedures). As shown in Fig. 7, SORD and NCL levels in adenomas and cancers tend to be higher in larger lesions (panel A). Furthermore, levels of SORD, S100A11, and SERPINB5 were more positively correlated with adenoma-associated mucosal pit patterns (IIs, IIIs, and IV) and histology (TA, TVA, VA) than with

cancer-associated pit pattern (type V) and histology (panels B and C, respectively) (3). When correlation analysis was extended to all the proteins that were significantly dysregulated in adenomas and/or cancers, PGM5 and G6PD levels displayed significant correlation with patient age, whereas SET, AKR1B1, and PPA2 abundance showed significant correlations with tumor size based on p values (supplemental Figs. S7). Furthermore, an array of proteins outlined in supplemental Fig. S8 significantly associated with tumor mucosal pit pattern, tumor histology and colon segment where tumor was localized (Table II).

SRM-based measurements of the selected 23 proteins were in general consistent with corresponding figures obtained with iTRAQ technology in our previous study in another adenoma series (34): ICC (A,1) = 0.65, p value = 0.00010, 95% CI = 0.353 - 0.828. We also calculated the Pearson's correlation coefficient between SRM and iTRAQ measurements using the average deconvoluted normal versus adenoma ratio for each protein (see Experimental Procedures for explanation of protein deconvoluted ratios). A Pearson's r -value of 0.82 showed a positive correlation between both data sets (Fig. 8A). However, few iTRAQ changes were not validated by SRM. The up-regulation of APEX1, RCN1 and SET observed with iTRAQ in adenomas were not confirmed by SRM, but they were detected in the adenocarcinomas analyzed by SRM (Fig. 3). ARF6 levels were up-regulated in the adenomas analyzed by iTRAQ and downregulated in the adenomas of the present study (Fig. 8A). To explore this discrepancy, we recalculated the protein abundance ratio of ARF6 in the iTRAQ adenomas—first using the sum of reporter ion channels from all confidently identified peptides (34), then with reporter ion intensity from the unique peptide employed for SRM quantification (supplemental Table S8)—and compared the resulting deconvoluted protein ratios with the one from the SRM study (Fig. 8B). As expected, ARF6 protein ratio based on all peptides confidently identified by iTRAQ indicated up-regulation in adenomas, whereas that based solely on the proteotypic peptide employed for SRM quantification showed a similar regulation trend with no significant change in expression from that of normal mucosa. This inconsistency in iTRAQ and SRM measurements for ARF6 might be attributed to biological variance inherent to patient samples. The adenoma-related down-regulation of ARF6 detected in the patient samples examined by SRM is most likely accurate because this protein is even more substantially downregulated in adenocarcinomas (Fig. 3 and Table II), and its expression is positively correlated with a more advanced (type V) mucosal pit pattern (Fig. 8C and Table II).

DISCUSSION

This study demonstrates that SRM technique can be used to quantify levels of multiple proteins that are putative markers of colorectal neoplasia across a relatively large set of tissue samples. In the assay, we developed, tissue protein extracts

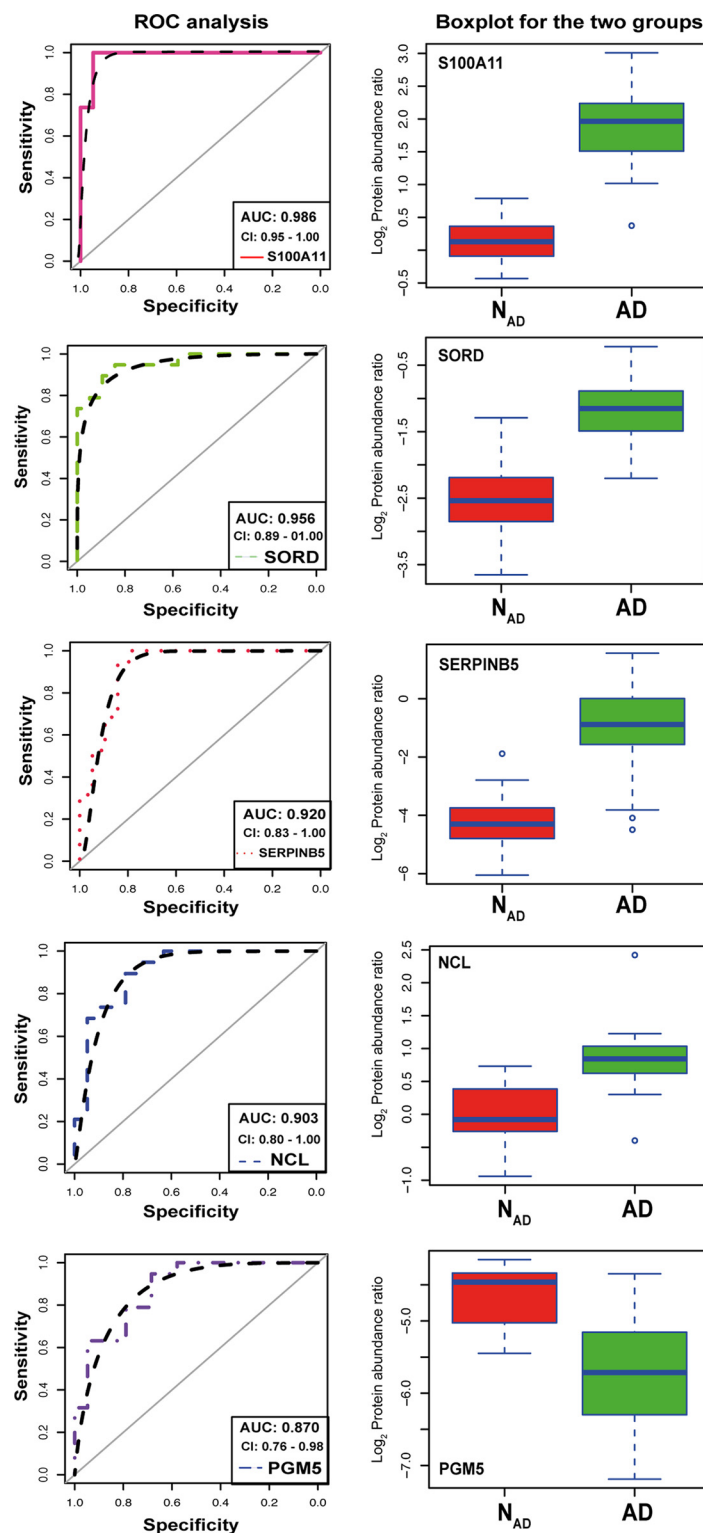


FIG. 4. **Predictive power of colorectal adenoma markers.** ROC analysis was performed for the five proteins most markedly dysregulated in adenomas (selected based on *p* values). ROC curves show the predictive power of each protein to distinguish between adenomas and normal mucosa (left panels). Protein intensities used for ROC analysis are shown in the box plot (right panels). N_{AD}; normal mucosa samples, AD: adenomas.

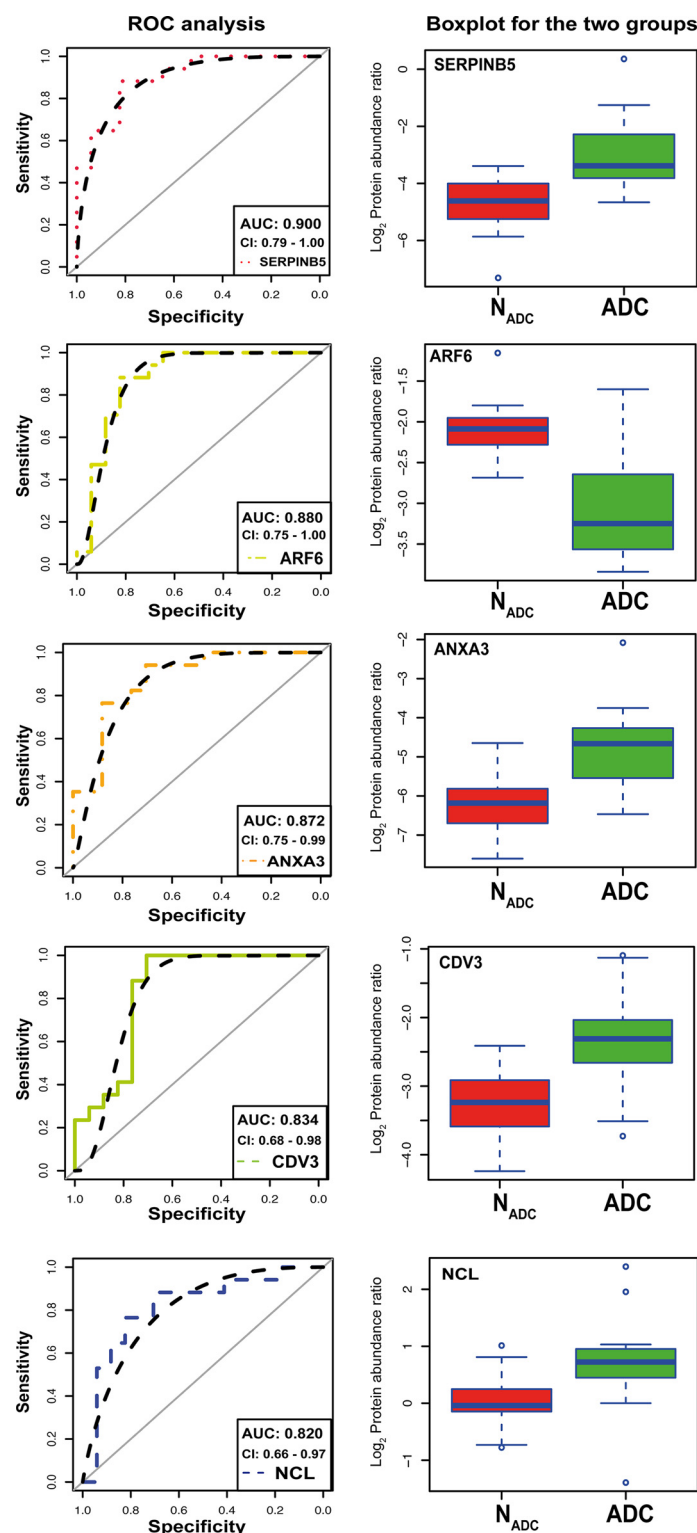


FIG. 5. **Predictive power of colorectal adenocarcinoma markers.** ROC analysis was performed for the five proteins most markedly dysregulated in adenocarcinomas (selected on the basis of *p* values). ROC curves show the power of each protein to distinguish between adenocarcinoma and normal mucosa (left panels). Protein intensities used for ROC analysis are shown in the box plot (right panels). N_{ADC}; normal mucosa samples, ADC: adenocarcinomas.

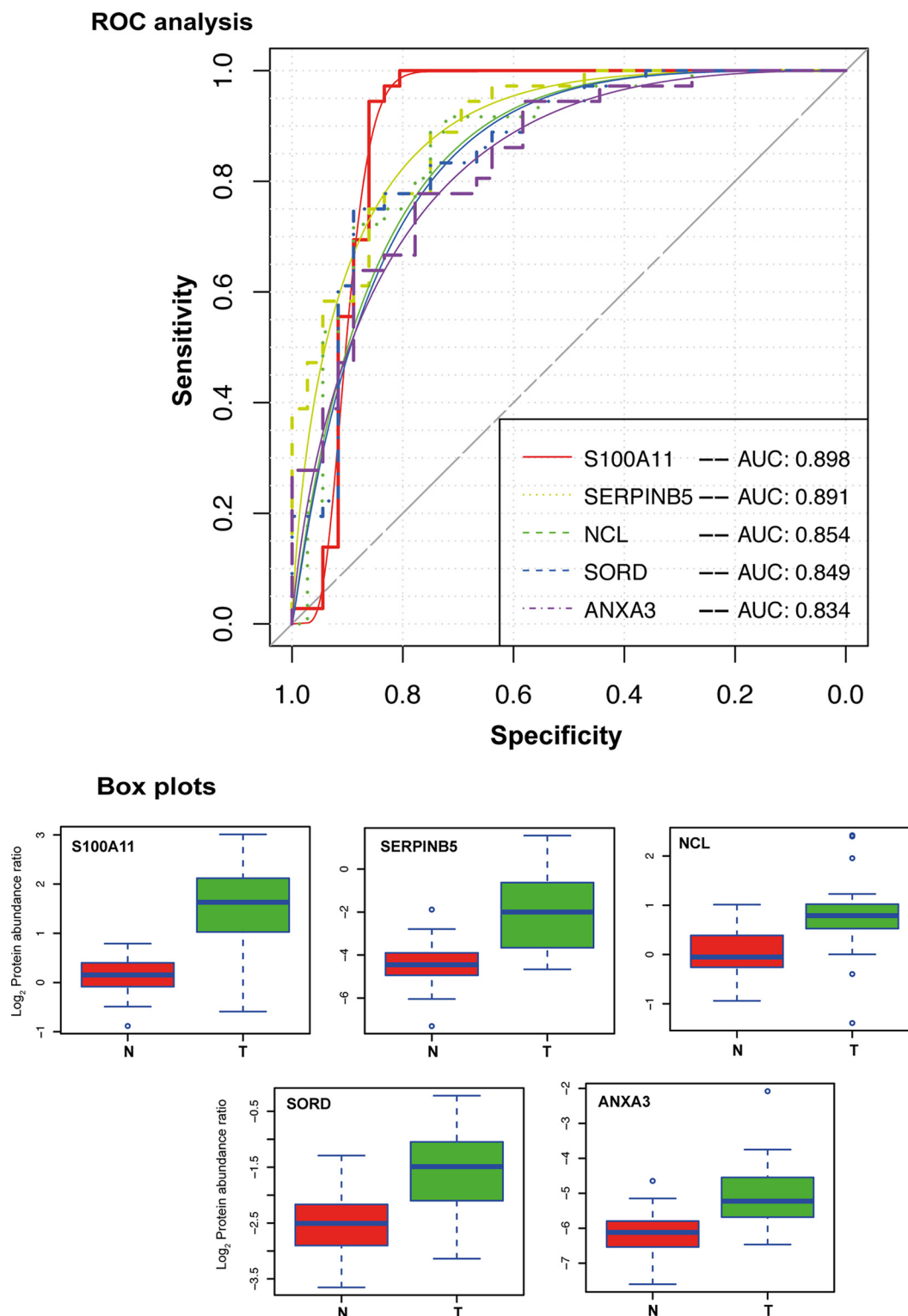


FIG. 6. **Potential markers for colorectal tumors (adenoma or adenocarcinoma).** ROC analysis was performed for the five proteins displaying the highest dysregulation in adenomas and adenocarcinomas (ranked according to *p* values). ROC curves show the power of each protein to distinguish between tumors and normal mucosa (*upper panel*). Protein intensities used for ROC analysis are shown in the box plot (*lower panels*). N; normal mucosa samples, T: adenomas and adenocarcinomas.

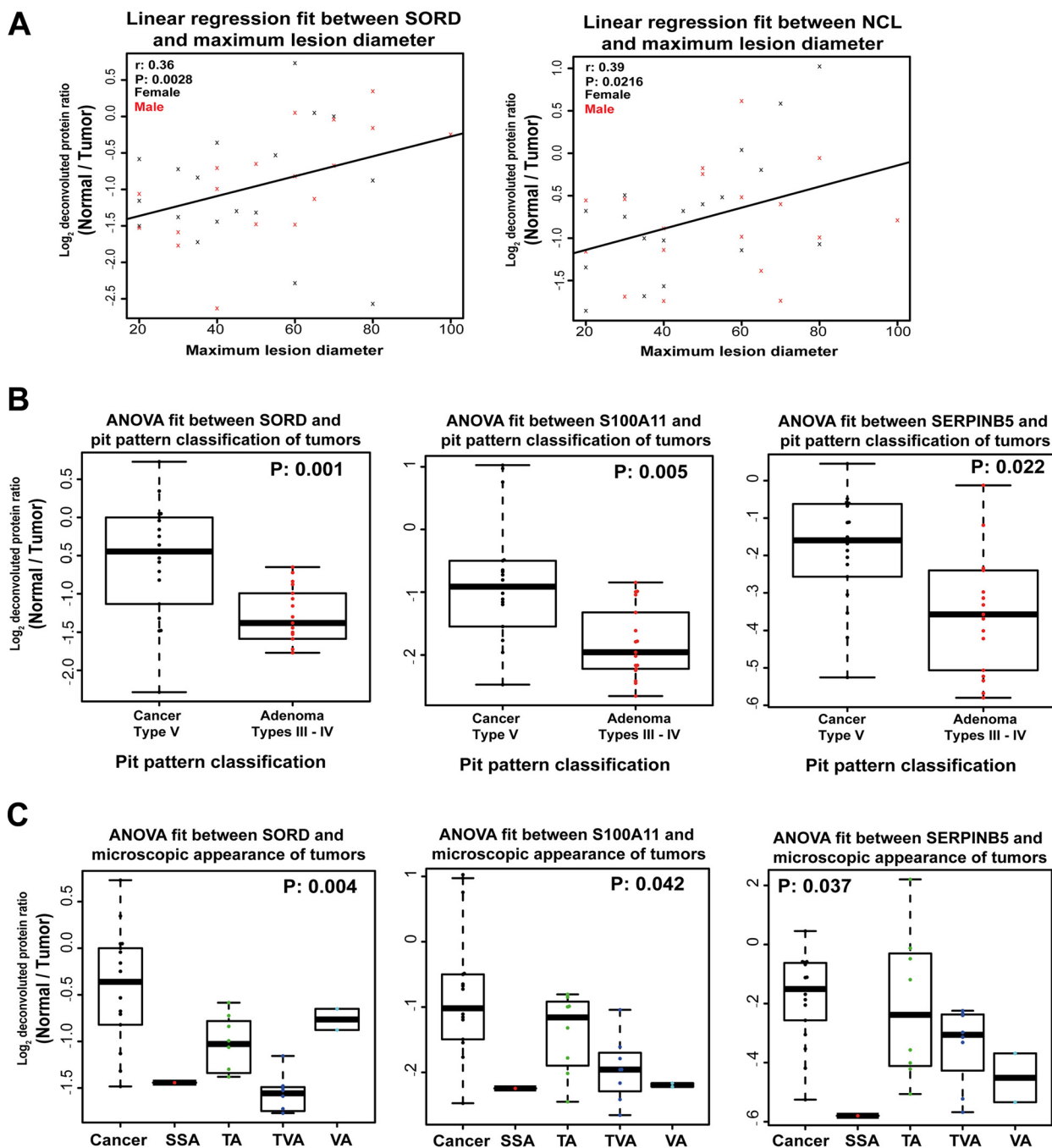


FIG. 7. Correlation between the five-marker protein panel and clinical features of each tumor. A, Linear fit shows the positive correlation between SORD or NCL abundance and lesion diameter. B, SORD, S100A11, and SERPINB5 were also positively associated with type V pit pattern in the neoplastic mucosa. C, Tumor expression of SORD, S100A11, and SERPINB5 was higher in adenoma subtypes than in adenocarcinomas (see Table I). Cancer - adenocarcinomas, SSA - sessile serrated adenoma, TA - tubular adenomas, TVA - tubulovillous adenomas, VA - villous adenomas.

could be analyzed without being subjected to sample fractionation or depletion of high-abundance proteins. Randomization and blocking were incorporated into the experimental design to maximize the chances of detecting true quantitative differences between normal and neoplastic (adenomas or carcinomas) colorectal tissues. Thanks to their

high reproducibility, our sample preparation and SRM measurement protocols (supplemental Figs. S2 and S3) can be readily adapted for use in various laboratories. Our panel of five proteins displaying tumor-associated dysregulation predicted the presence of both precancerous and invasive colorectal tumors.

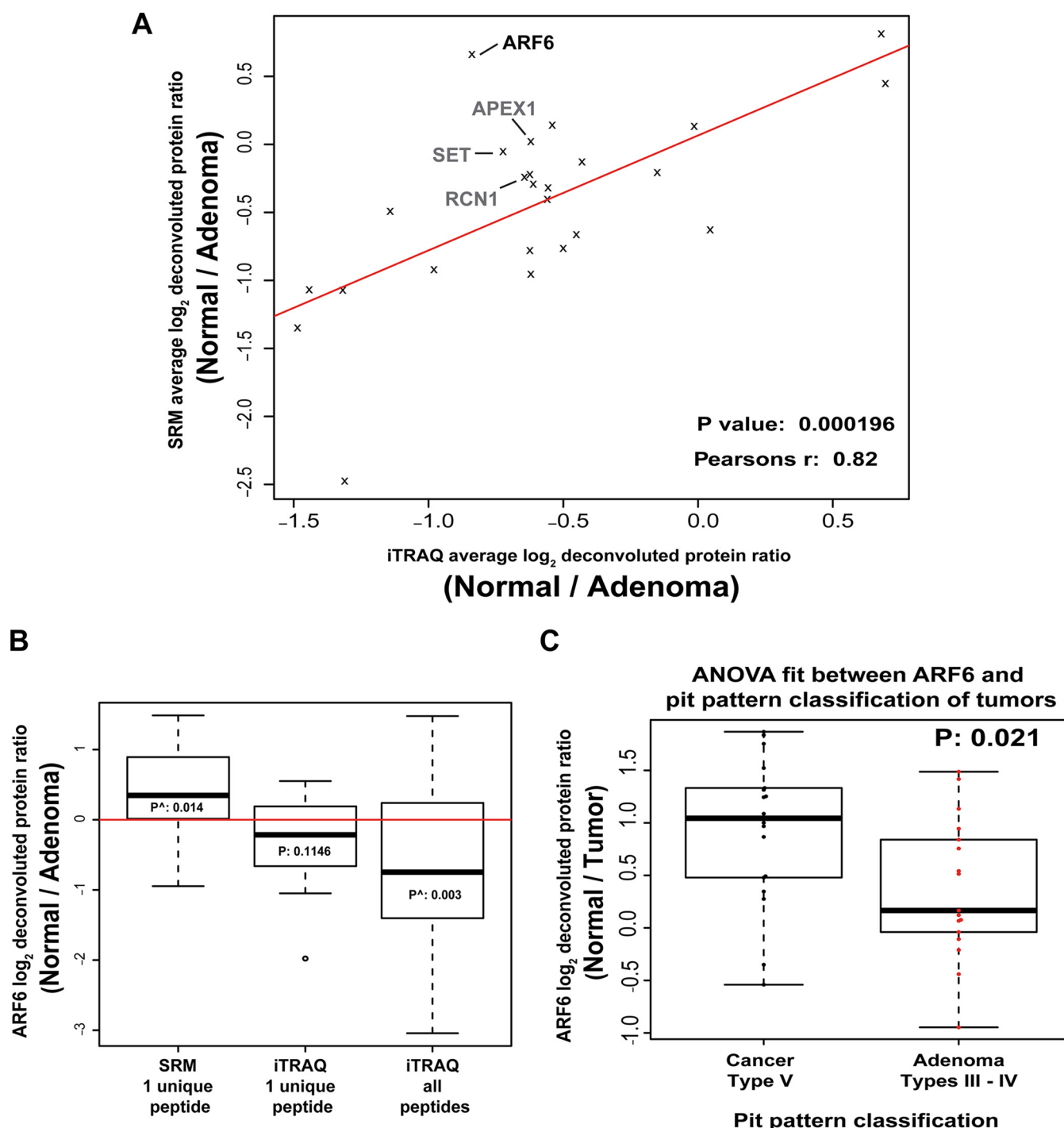


FIG. 8. Correlation between SRM measurements of target proteins in adenomas in the present study and corresponding iTRAQ measurements from our previous study (34). A, Analysis of the average deconvoluted protein ratios (normal mucosa/adenoma) obtained with the two methods shows fairly good correlation between the two sets of measurements. p value = 0.000196, Pearson's r = 0.82. B, Deconvoluted \log_2 ratio (normal/adenoma) of ARF6 in tumors investigated with iTRAQ and with SRM. ARF protein ratio in matched normal and adenoma samples was quantified based on (1) the intensity of the unique peptide consistently quantified in samples in the SRM study (SRM 1 unique peptide), (2) the intensity of this same peptide measured in samples from the iTRAQ group (iTRAQ 1 unique peptide), and (3) the intensity of all confidently identified peptides measured with iTRAQ 8-plex method (iTRAQ all peptides). Protein ratio in adenomas and matched normal samples was compared with a paired t test and $^{\wedge}$ indicates a significant p value (C) ANOVA of fit between ARF6 abundance and tumor pit pattern.

The fact that our SRM assays failed to quantify 12 of the 40 proteins we targeted might be related to the detection limit we set for low-abundance proteins in our unfractionated, complex samples (54, 55), because it was higher than that used for fractionated peptide digests in our previous study (34). It is also possible that the endogenous proteotypic peptides we selected to represent these proteins had suboptimal MS signal responses. Despite these shortcomings, our experience in this study clearly highlighted the advantages of targeted proteomics platforms over the shotgun (or discovery) MS methods (e.g. iTRAQ) used in our previous study (34)), which include higher quantitative sensitivity, improved reproducibility, and multiplex capabilities (56). The emerging targeted MS methods, Sequential Window Acquisition of all Theoretical Spectra (SWATH) and Parallel Reaction Monitoring, may well provide even more comprehensive proteome profiles and further expand the degree of multiplexing achieved with SRM (14–16).

The proteotypic peptides targeted by SRM are selected *a priori* to facilitate differentiation between protein isoforms or proteins with shared peptides. In addition, the targeted peptide and fragment ions are reproducibly monitored in all samples, which improves selectivity (14, 17, 57). In our previous shotgun MS study (34), for example, adenoma-associated expression changes in LDH, ANP32, and PPA could not be attributed to a specific isoform of the enzymes because similar peptides were used for the relative quantification of both isoforms (LDHA *versus* LDHB, ANP32A *versus* ANP32B and PPA1 *versus* PPA2). In the present study, we chose peptides that were unique to each protein/isoform and thus obtained isoform-specific information on the altered abundance of these enzymes in colorectal tumors (Table II and [supplemental Fig. 9](#)).

On the whole, SRM-based measurements of the proteins of interest displayed good agreement and correlation with corresponding figures obtained with an 8-plex iTRAQ approach (Fig. 8A). The main outlier was ARF6, which displayed reduced adenoma-related expression in the present study whereas iTRAQ analysis had revealed increased levels in the precancerous tumors. This inconsistency mainly reflects heterogeneity in ARF6 expression across the patient tumors examined. It is not wholly related to the use of one unique peptide for SRM quantification of ARF6 because the reported expression trend for this protein in the iTRAQ cohort did not change when protein intensity was recalculated based on only this peptide (Fig. 8B). Moreover, SRM measurements based on the unique peptide for ARF6 demonstrated adenoma-related down-regulation of this protein in an independent cohort, which was supported by the positive correlation with more severe mucosal changes (Fig. 8C) and progression to cancer (Fig. 3).

Compression of signal intensity ratios secondary to coisolation and fragmentation of tagged precursors is a known drawback of iTRAQ labeling (58–60). Others include under-

sampling because of stochastic selection of precursor ions and intensity-biased selection of precursor ions (59, 60). All of these phenomena can cause variability across samples in protein identification and quantification (17, 61), underscoring once again the need for more sensitive and reproducible methods such as SRM for the validation of potential biomarker candidates.

Regular colonoscopic screening is recommended for individuals with age-related risk factors for CRC. The reported sensitivity of this approach ranges from 89% to 98% for the detection of adenomas 10 mm or larger and from 75% to 93% for those measuring 6 mm or more (62). Colonoscopy rarely misses adenomas 10 mm or larger (2.1%), but miss rates increase with smaller adenomas (63). In addition, the efficacy of colonoscopic screening programs is often reduced by low patient participation rates (64, 65). Efforts have thus been made to find cheaper, noninvasive alternatives that can detect both CRC and advanced precancerous lesions (*i.e.* >10 mm large adenomas) with high sensitivity, thereby improving patient compliance and clinical outcomes. The past decade has witnessed the discovery of a number of potential biomarkers for the diagnosis of colorectal neoplasms (6, 66, 67), but the only non-invasive screening methods for CRC approved thus far involve immunoassays for fecal hemoglobin (68). Studies are underway to assess the efficacy of a more recently developed multi-target stool DNA test, which combines hemoglobin detection with quantitative molecular assays for *KRAS* mutation, β -actin, and aberrant *NDRG4* and *BMP3* methylation (68). In a large cross-sectional validation study on asymptomatic individuals at average risk for CRC, this new approach detected significantly more cancers than the fecal immuno test, but it also yielded more false-positive results. Its sensitivity for the detection of advanced adenomas—admittedly mediocre (42.4%)—was clearly higher than that of the fecal immuno test (23.8%).

Unlike DNA, protein-based markers provide an accurate representation of the functional cellular conditions being investigated and are fundamental for the phenotypic diagnosis of a disease. For example, mutation of a gene might affect different domains of the protein it encodes, resulting in diverse phenotypes. Furthermore, DNA aberrations generated by double-strand breaks can be amplified into a variety of protein modification events. However, assays for measuring protein levels in fecal samples are not as well established as stool DNA assays despite reports that stool samples from patients with CRC can be distinguished from those of controls based on levels of several tumor-derived proteins, including pyruvate kinase M2 (69, 70), carcinoembryonic antigen (CEA) (70), secreted clusterin isoform (71), and minichromosome maintenance proteins (72). The application of proteomics to the analysis of stool samples has been hindered by the biological processes involving partial digestion and decay of proteins throughout the gastrointestinal tract, as well as by

technical challenges related to sample preservation and processing.

Disease-specific eukaryotic protein markers enter the extracellular space via conventional and/or unconventional secretory pathways (73). Proteins with signal peptides are usually secreted through conventional pathways involving the endoplasmic reticulum-Golgi system. Unconventional secretory pathways include (1) translocation across the plasma membrane, (2) lysosome-dependent pathways, and (3) extracellular vesicles-dependent secretion (74). We examined probable secretion pathways for the 23 candidate proteins.

First, the SignalP server 4.0 (75) was used to identify protein(s) that possessed a signal peptide and were actively secreted via the classical pathway, whereas the SecretomeP server 2.0a (76) was used to predict protein secretion via non-classical pathways for proteins with or without signal peptides (Table II). At a Discrimination score (D-score) cutoff ≥ 0.45 , NGAL, OLFM4, RCN1, and REG4 are most likely secreted via the classical pathways. Four additional proteins (APEX1, ARF6, S100A11, and KCTD12) were predicted by SecretomeP (Neural Network output or NN-score ≥ 0.60) to undergo non-classical secretion.

Second, because non-classical secretion of proteins can also occur via their externalization in extra-cellular vesicles (e.g. microvesicles, exosomes, and apoptotic bodies), we analyzed our candidates using Vesiclepedia (77), a manually curated compendium of molecules identified in different classes of extracellular vesicles. Surprisingly, all our candidate proteins have been detected in microvesicles and exosomes from colorectal cells (Table II). Tumor cells can also be shed into the blood. Because extracellular vesicles and circulating tumor cells can be robustly and reproducibly detected in blood and other body fluids, it seems reasonable to expect that our candidate protein biomarkers are potentially detectable in blood (78). This conclusion is supported by the fact that 61% of our target proteins (Table II) were found in the high-confidence human plasma proteome reference set reported by Farrah and colleagues (79), and 96% (all except OLFM4) were present in the largest high-confidence plasma proteome data set available to date, that generated by Keshishian and colleagues using a novel plasma-based biomarker discovery workflow that increased sample throughput and improved the depth of proteome identification and relative quantification (80). Leakage into the blood stream of proteins from early-stage adenomas is probably unlikely given the small size of these tumors and the barriers the proteins would have to circumvent prior to reaching the blood. There is a higher chance of detecting proteins from advanced adenomas or adenocarcinomas in the plasma/serum because of leakage, shedding or protein secretion. For example, studies on one of our candidate biomarkers, SERPINB5, confirmed that levels of this protein are increased in serum from patients with colorectal adenomas or adenocarcinomas, as compared with healthy individuals (81).

Certain molecular changes associated with tumor occurrence and progression can be observed in multiple types of cancer and even in certain chronic inflammatory diseases. Their detection in body fluids might thus be fairly nonspecific. We examined 23 of the 25 protein targets listed in Table II for possible associations with other cancer types and/or with inflammatory bowel diseases. A review of Polanski and Anderson's list of 1261 proteins reported to be differentially expressed in human cancers (82) indicated that eight of these proteins (35% of our protein list) were associated with cancers other than CRC (Table II). FunDO (45) analysis of the genes encoding these 23 candidates revealed six additional proteins that were associated with multiple cancers, but not with inflammatory bowel diseases (supplemental Fig. 10). These findings indicate that our proteins of interest are not associated with inflammation of the digestive tract, however fourteen (56%) of them have links to multiple cancers.

Carcinoembryonic antigen is the only protein currently used as a circulating biomarker for CRC recurrence and metastasis (83). The proteins that were consistently up-regulated in our colorectal tumors and that distinguished adenoma and/or adenocarcinomas from normal mucosa with appreciable sensitivity and specificity (Figs. 4 to 6) are promising candidates for use in blood-based assays for the detection of these tumors. Some of these proteins have been already investigated in other proteomics studies as prognostic serum biomarkers in patients with CRC. Elevated serum levels of REG4 and of total LDH, for example, have both proved to be significant indicators of liver metastasis in this setting (84, 85), whereas increased serum levels of LCN2 are reportedly associated with higher neoplastic volume and disease recurrence (86).

Our data suggest that the presence in body fluids of S100A11, SERPINB5, NCL, SORD, and ANXA3 might be an indication for colonoscopy, although this five-marker signature cannot distinguish whether the patient is likely to be harboring an adenoma or an adenocarcinoma. Three of these markers have also been reported in blood-based validation studies on CRC. For instance, a nine-protein multiplex serum array biochip that includes S100A11 and CEA showed high diagnostic potential for colon cancer (87). Serum levels of SERPINB5 have also been found to be elevated in patients with colorectal adenoma or CRC (81), and ANXA3 was recently validated as a potential diagnostic and therapeutic serum marker for CRC (88).

The substances secreted in detectable amounts by CRCs also include products of tumor metabolism. The analysis of volatile organic compounds is an emerging field in the search for noninvasive metabolite tumor-marker profiles (89). Electronic-nose technology and canine and *Caenorhabditis elegans* scent detection have been used to analyze the presence of these compounds in breath, urine, and stool samples (89–91) from CRC patients. The reported sensitivity and specificity of these approaches in distinguishing between individ-

uals with and without CRC or advanced adenomas are comparable to those of conventional colonoscopy (>90%) (90, 91). In the present study, we found significantly increased levels of metabolic enzymes in the adenomas (SORD, LDHA) and adenocarcinomas (SORD, G6PD, LDHA) we examined. Metabolite products from reactions catalyzed by these enzymes and associated pathway enzymes are potential indicators of tumor occurrence. Interestingly, lactate levels in the adenomas we examined in (34) were significantly higher in adenomas than in healthy control tissue, suggesting that aerobic glycolysis occurs in benign colorectal tumors.

The potential translational impact of our findings on the future clinical management of patients with early colorectal lesions is obvious. However, they can also be exploited to improve our understanding of the molecular mechanisms underlying colorectal tumorigenesis. To this end, we probed the five proteins that most effectively discriminated between normal mucosa and colorectal tumors (Fig. 6) for functional relations with the rest of the proteome. Using the GeneMANIA tool, we queried association data on protein and genetic interactions, pathways, coexpression, colocalization, and protein domain similarity. As shown in [supplemental Fig. S11](#), nucleolin (NCL) was the signature protein with the highest number of interactions with cancer-related proteins. It directly associates with TERT and TOP1, which are key players in oncogenesis (92, 93). TERT mRNA and protein expression occurs early in *in vivo* colorectal tumorigenesis, beginning in the early premalignant phase and increasing gradually during progression to invasive stages (94). As for TOP1, inhibitors of this enzyme are used as first-line chemotherapeutics for CRC (93). The calcium binding protein S100A11 is best described as a dual mediator with effects that depend on its localization and interaction partners. It therefore functions as a tumor suppressor in some cancers (e.g. those of the breast) and as an oncogene in others, including CRC. It interacts directly with NCL, S100B, ANXA1, and UBE2F (95–98) ([supplemental Fig. S11](#)). The peptidase inhibitor, SERPINB5, has also been reported to exert tumor-promoting as well as tumor-suppressor effects. Recent protein biomarker studies have revealed that elevated SERPINB5 expression in CRCs correlates with high CEA levels and poor prognosis (99). SORD was found to physically interact with three proteins: AKR1B1, which—like SORD—is involved in the polyol pathway; GCLM, and C16orf13 (98). The roles of GCLM and C16orf13 in tumorigenesis are unclear. Annexin A3 is a calcium-dependent phospholipid-binding protein with potential roles in maintaining cell proliferation and invasion (88). It physically interacts with ANXA11 (98).

In conclusion, SRM allows reproducible, simultaneous detection and quantification of multiple proteins in numerous samples. Protein abundance changes in adenomas relative to the normal colorectal mucosa are unambiguously indicative of phenotypic alterations present in early stage tumors. Our

study is limited in regard to the validation of these potential biomarkers in serum/plasma of patients undergoing colonoscopic screening, and patients diagnosed with adenomas and/or CRC. Future work should assess levels of the proteins we verified in colorectal tissue samples in blood samples collected from patients. Efforts of this type will undoubtedly reveal additional, related proteins or metabolites with complementary diagnostic attributes that might also prove useful in precolonoscopic, non-invasive tests for the early diagnosis of colorectal tumors.

Additional Information—The MS/MS data from this study have been submitted to the PeptideAtlas SRM Experiment Library (PASSEL) (<http://www.peptideatlas.org/passel/>). SRM data are available at PeptideAtlas SRM Experiment Library (PASSEL) and can be accessed using the identifier PASS00900.

Acknowledgments—We thank Bernd Roschitzki, Yansheng Liu, Ruedi Aebersold, and Joseph Jiricny for their support and productive discussions and Marian Everett Kent and Owen Howard for editing the manuscript.

* This work was supported by a grant from the Swiss National Science Foundation (Grant No. 31003A_141225).

§ This article contains [supplemental material](#).

** To whom correspondence should be addressed: Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Tel.: 0041 44 635 3472; Fax 0041 44 635 3484; E-mail: marra@imcr.uzh.ch.

‡‡ Present address: Institute of Molecular Systems Biology, ETH Zurich, Switzerland.

The authors declare that they have no competing interests.

REFERENCES

1. Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA* **63**, 11–30
2. Siegel, R., Desantis, C., and Jemal, A. (2014) Colorectal cancer statistics, 2014. *CA* **64**, 104–117
3. Participants in the Paris Workshop . (2003) The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon. *Gastrointest. Endosc.* **58**, S3–S43
4. Cattaneo, E., Baudis, M., Buffoli, F., Bianco, M. A., Zorzi, F., and Marra, G. (2010) in *Pre-Invasive Disease: Pathogenesis and Clinical Management* (Springer New York, New York, NY), pp 369–394
5. Fearon, E. R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767
6. de Wit, M., Fijneman, R. J., Verheul, H. M., Meijer, G. A., and Jimenez, C. R. (2013) Proteomics in colorectal cancer translational research: Biomarker discovery for clinical applications. *Clin. Biochem.* **46**, 466–479
7. Cunningham, D., Atkin, W., Lenz, H. J., Lynch, H. T., Minsky, B., Nordlinger, B., and Starling, N. (2010) Colorectal cancer. *Lancet* **375**, 1030–1047
8. Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. (2010) Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J. Proteomics* **73**, 1873–1895
9. Luo, Y., Wang, L., and Wang, J. (2013) Developing proteomics-based biomarkers for colorectal neoplasms for clinical practice: Opportunities and challenges. *Prot. Clin. Clin. Appl.* **7**, 30–41
10. Ang, C. S., Phung, J., and Nice, E. C. (2011) The discovery and validation of colorectal cancer biomarkers. *Biomed. Chromatogr.* **25**, 82–99
11. Song, K., Fendrick, A. M., and Ladabaum, U. (2004) Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. *Gastroenterology* **126**, 1270–1279
12. Davies, R. J., Miller, R., and Coleman, N. (2005) Colorectal cancer screening: prospects for molecular stool analysis. *Nat. Rev. Cancer* **5**, 199–209

13. Liu, Y., Hüttenhain, R., Collins, B., and Aebersold, R. (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Rev. Mol. Diagn.* **13**, 811–825
14. Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., and Aebersold, R. (2010) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **7**, 43–46
15. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **11**, O111.016717
16. Domon, B., and Gallien, S. (2015) Recent advances in targeted proteomics for clinical applications. *Prot. Clin. Appl.* **9**, 423–431
17. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222
18. Malmström, L., Malmström, J., Selevsek, N., Rosenberger, G., and Aebersold, R. (2012) Automated Workflow for Large-Scale Selected Reaction Monitoring Experiments. *J. Proteome Res.* **11**, 1644–1653
19. Parker, C. E., and Borchers, C. H. (2014) Mass spectrometry based biomarker discovery, verification, and validation – Quality assurance and control of protein biomarker assays. *Mol. Oncol.* **8**, 840–858
20. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566
21. Füzéry, A. K., Levin, J., Chan, M. M., and Chan, D. W. (2013) Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin. Proteomics* **10**, 13
22. Baylin, S. B., and Ohm, J. E. (2006) Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* **6**, 107–116
23. Hüttenhain, R., Malmström, J., Picotti, P., and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* **13**, 518–525
24. Lopez, M. F., Kuppasamy, R., Sarracino, D. A., Prakash, A., Athanas, M., Krastins, B., Rezai, T., Sutton, J. N., Peterman, S., and Nicolaides, K. (2011) Mass Spectrometric Discovery and Selective Reaction Monitoring (SRM) of Putative Protein Biomarker Candidates in First Trimester Trisomy 21 Maternal Serum. *J. Proteome Res.* **10**, 133–142
25. Muraoka, S., Kume, H., Watanabe, S., Adachi, J., Kuwano, M., Sato, M., Kawasaki, N., Kodera, Y., Ishitobi, M., Inaji, H., Miyamoto, Y., Kato, K., and Tomonaga, T. (2012) Strategy for SRM-based Verification of Biomarker Candidates Discovered by iTRAQ Method in Limited Breast Cancer Tissue Samples. *J. Proteome Res.* **11**, 4201–4210
26. Selevsek, N., Matondo, M., Sanchez Carbayo, M., Aebersold, R., and Domon, B. (2011) Systematic quantification of peptides/proteins in urine using selected reaction monitoring. *Proteomics* **11**, 1135–1147
27. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Huttenhain, R., Koomen, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R., Anderson, L., Bethem, R., Blonder, J., Boja, E., Botelho, J., Boyne, M., Bradshaw, R. A., Burlingame, A. L., Chan, D., Keshishian, H., Kuhn, E., Kinsinger, C., Lee, J. S., Lee, S. W., Moritz, R., Oses-Prieto, J., Rifai, N., Ritchie, J., Rodriguez, H., Srinivas, P. R., Townsend, R. R., Van Eyk, J., Whiteley, G., Wiita, A., and Weintraub, S. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13**, 907–917
28. Cretu, D., Liang, K., Saraon, P., Batruch, I., Diamandis, E. P., and Chandran, V. (2015) Quantitative tandem mass-spectrometry of skin tissue reveals putative psoriatic arthritis biomarkers. *Clin. Proteomics* **12**, 1
29. Proc, J. L., Kuzyk, M. A., Hardie, D. B., Yang, J., Smith, D. S., Jackson, A. M., Parker, C. E., and Borchers, C. H. (2010) A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J. Proteome Res.* **9**, 5422–5437
30. Cretu, D., Prassas, I., Saraon, P., Batruch, I., Gandhi, R., Diamandis, E. P., and Chandran, V. (2014) Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry. *Clin. Proteomics* **11**, 27
31. Cerciello, F., Choi, M., Nicastrì, A., Bausch-Fluck, D., Ziegler, A., Vitek, O., Felley-Bosco, E., Stahel, R., Aebersold, R., and Wollscheid, B. (2013) Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. *Clin. Proteomics* **10**, 16
32. Kim, K., Kim, S. J., Yu, H. G., Yu, J., Park, K. S., Jang, I. J., and Kim, Y. (2010) Verification of Biomarkers for Diabetic Retinopathy by Multiple Reaction Monitoring. *J. Proteome Res.* **9**, 689–699
33. Meng, Z., and Veenstra, T. D. (2011) Targeted mass spectrometry approaches for protein biomarker verification. *J. Proteomics* **74**, 2650–2659
34. Uzozie, A., Nanni, P., Staiano, T., Grossmann, J., Barkow-Oesterreicher, S., Shay, J. W., Tiwari, A., Buffoli, F., Laczkó, E., and Marra, G. (2014) Sorbitol dehydrogenase overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: a quantitative proteomics study. *Mol. Cell. Proteomics* **13**, 1198–1218
35. Glatter, T., Ludwig, C., Ahme, E., Aebersold, R., Heck, A. J., and Schmidt, A. (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11**, 5145–5156
36. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., and Rinner, O. (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111–1121
37. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
38. Maclean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., Whiteaker, J. R., Paulovich, A. G., Carr, S. A., and MacCoss, M. J. (2010) Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal. Chem.* **82**, 10116–10124
39. Oberg, A. L., and Vitek, O. (2009) Statistical design of quantitative mass spectrometry-based proteomic experiments. *J. Proteome Res.* **8**, 2144–2156
40. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M. Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8**, 430–435
41. Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526
42. Chang, C. Y., Picotti, P., Huttenhain, R., Heinzelmann-Schwarz, V., Joivanovic, M., Aebersold, R., and Vitek, O. (2012) Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol. Cell. Proteomics* **11**, M111.014662/M111.014662
43. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868
44. Müller, R., and Büttner, P. (1994) A critical discussion of intraclass correlation coefficients. *Statist. Med.* **13**, 2465–2476
45. Exploring genes using functional disease ontology annotations exploring genes using functional disease ontology annotations. *projects.bioinformatics.northwestern.edu*
46. Osborne, J. D., Flatow, J., Holko, M., Lin, S. M., Kibbe, W. A., Zhu, L. J., Dania, M. I., Feng, G., and Chisholm, R. L. (2009) Annotating the human genome with Disease Ontology. *BMC Genomics* **10**, 1–21
47. Du, P., Feng, G., Flatow, J., Song, J., Holko, M., Kibbe, W. A., and Lin, S. M. (2009) From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. *Bioinformatics* **25**, i63–i68
48. Falcon, S., and Gentleman, R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258
49. Cattaneo, E., Laczkó, E., Buffoli, F., Zorzi, F., Bianco, M. A., Menigatti, M., Bartosova, Z., Haider, R., Helmchen, B., Sabates-Bellver, J., Tiwari, A., Jiricny, J., and Marra, G. (2011) Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid). *EMBO Mol. Med.* **3**, 334–347

50. Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Edes, J., Loewenich, S. N., and Aebersold, R. (2006) The PeptideAtlas project. *Nucleic Acids Res.* **34**, D655–D658
51. Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265–273
52. Brönstrup, M. (2004) Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Rev. Proteomics* **1**, 503–512
53. Ebhardt, H. A., Sabido, E., Hüttenhain, R., Collins, B., and Aebersold, R. (2012) Range of protein detection by selected/multiple reaction monitoring mass spectrometry in an unfractionated human cell culture lysate. *Proteomics* **12**, 1185–1193
54. Surinova, S., Schiess, R., Hüttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011) On the development of plasma protein biomarkers. *J. Proteome Res.* **10**, 5–16
55. Kulasingam, V., Pavlou, M. P., and Diamandis, E. P. (2010) Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer* **10**, 371–378
56. Dornon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28**, 710–721
57. Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomäki, P., de la Chapelle, A., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1996) Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat. Med.* **2**, 169–174
58. Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009) iTRAQ Underestimation in simple and complex mixtures: “The Good, the Bad and the Ugly.” *J. Proteome Res.* **8**, 5347–5355
59. Savitski, M. M., Mathieson, T., Zinn, N., Sweetman, G., Doce, C., Becher, I., Pachi, F., Kuster, B., and Bantscheff, M. (2013) Measuring and for managing ratio compression accurate iTRAQ/TMT quantification. *J. Proteome Res.* **12**, 3586–3598
60. Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V., and Lilley, K. S. (2010) Addressing accuracy and precision issues in iTRAQ quantitation. *Mol. Cell. Proteomics* **9**, 1885–1897
61. Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007) Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* **6**, 821–827
62. Lin, J. S., Piper, M. A., Perdue, L. A., Rutter, C. M., Webber, E. M., O'Connor, E., Smith, N., and Whitlock, E. P. (2016) Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* **315**, 2576–2594
63. van Rijn, J. C., Reitsma, J. B., Stoker, J., Bossuyt, P. M., van Deventer, S. J., and Dekker, E. (2006) Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am. J. Gastroenterol.* **101**, 343–350
64. Garborg, K., Holme Ø Loberg, M., Kalager, M., Adami, H. O., and Bretthauer, M. (2013) Current status of screening for colorectal cancer. *Ann. Oncol.* **24**, 1963–1972
65. Levin, B., Lieberman, D. A., McFarland, B., Smith, R. A., Brooks, D., Andrews, K. S., Dash, C., Giardiello, F. M., Glick, S., Levin, T. R., Pickhardt, P., Rex, D. K., Thorson, A., Winawer, S. J., for the American Cancer Society Colorectal Cancer Advisory Group the Multi-Society Task Force, U. S., and the American College of Radiology Colon Cancer Committee. (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA* **58**, 130–160
66. Shin, J., Kim, H. J., Kim, G., Song, M., Woo, S. J., Lee, S. T., Kim, H., and Lee, C. (2014) Discovery of melanotransferrin as a serological marker of colorectal cancer by secretome analysis and quantitative proteomics. *J. Proteome Res.* **13**, 4919–4931
67. Kume, H., Muraoka, S., Kuga, T., Adachi, J., Narumi, R., Watanabe, S., Kuwano, M., Koda, Y., Matsushita, K., Fukuoka, J., Masuda, T., Ishihama, Y., Matsubara, H., Nomura, F., and Tomonaga, T. (2014) Discovery of colorectal cancer biomarker candidates by membrane proteomic analysis and subsequent verification using selected reaction monitoring (SRM) and tissue microarray (TMA) analysis. *Mol. Cell. Proteomics* **13**, 1471–1484
68. Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Levin, T. R., Lavin, P., Lidgard, G. P., Ahlquist, D. A., and Berger, B. M. (2014) Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* **370**, 1287–1297
69. Hathurusinghe, H. R., Goonetilake, K. S., and Siriwardena, A. K. (2007) Current status of tumor M2 pyruvate kinase (Tumor M2-PK) as a biomarker of gastrointestinal malignancy. *Ann. Surg. Oncol.* **14**, 2714–2720
70. Kim, Y., Lee, S., Park, S., Jeon, H., Lee, W., Kim, J. K., Cho, M., Kim, M., Lim, J., Kang, C. S., and Han, K. (2003) Gastrointestinal tract cancer screening using fecal carcinoembryonic antigen. *Ann. Clin. Lab. Sci.* **33**, 32–38
71. Pucci, S., Bonanno, E., Sesti, F., Mazzarelli, P., Mauriello, A., Ricci, F., Zoccai, G. B., Rulli, F., Galatà, G., and Spagnoli, L. G. (2009) Clusterin in stool: a new biomarker for colon cancer screening? *Am. J. Gastroenterology* **104**, 2807–2815
72. Davies, R. J., Freeman, A., Morris, L. S., Bingham, S., Dilworth, S., Scott, I., Laskey, R. A., Miller, R., and Coleman, N. (2002) Analysis of minichromosome maintenance proteins as a novel method for detection of colorectal cancer in stool. *Lancet* **359**, 1917–1919
73. Nickel, W., and Rabouille, C. (2009) Mechanisms of regulated unconventional protein secretion. *Nat. Rev. Mol. Cell Biol.* **10**, 148–155
74. de Wit, M., Kant, H., Piersma, S. R., Pham, T. V., Mongera, S., van Berkel, M. P., Boven, E., Pontén, F., Meijer, G. A., Jimenez, C. R., and Fijneman, R. J. (2014) Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. *J. Proteomics* **99**, 26–39
75. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795
76. Bendtsen, J. D., Jensen, L. J., Blom, N., von Heijne, G., and Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **17**, 349–356
77. Kalra, H., Simpson, R. J., Ji, H., Aikawa, E., Altevogt, P., Askenase, P., Bond, V. C., Borras, F. E., Breakefield, X., Budnik, V., Buzas, E., Camussi, G., Clayton, A., Cocucci, E., Falcon-Perez, J. M., Gabrielson, S., Gho, Y. S., Gupta, D., Harsha, H. C., Hendrix, A., Hill, A. F., Inal, J. M., Jenster, G., Kraemer-Albers, E. M., Lim, S. K., Llorente, A., Lotvall, J., Marcilla, A., Mincheva-Nilsson, L., Nazarenko, I., Nieuwland, R., Nolte-t Hoen, E. N., Pandey, A., Patel, T., Piper, M. G., Pluchino, S., Prasad, T. S., Rajendran, L., Raposo, G., Record, M., Reid, G. E., Sanchez-Madrid, F., Schiffelers, R. M., Siljander, P., Stensballe, A., Stoorvogel, W., Taylor, D., Thery, C., Valadi, H., van Balkom, B. W., Vazquez, J., Vidal, M., Wauben, M. H. M., Yanez-Mo, M., Zoeller, M., and Mathivanan, S. (2012) Vesiclepedia: A compendium for extracellular vesicles with continuous community annotation. *PLoS Biol.* **10**(12): e1001450
78. Yap, T. A., Lorente, D., Omlin, A., Olmos, D., and de Bono, J. S. (2014) Circulating tumor cells: a multifunctional biomarker. *Clin. Cancer Res.* **20**, 2553–2568
79. Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmstrom, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L., and Aebersold, R. (2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **10**, M110.006353
80. Keshishian, H., Burgess, M. W., Gillette, M. A., Mertins, P., Clauser, K. R., Mani, D. R., Kuhn, E. W., Farrell, L. A., Gerszten, R. E., and Carr, S. A. (2015) Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. *Mol. Cell. Proteomics* **14**, 2375–2393
81. Wild, N., Andres, H., Rollinger, W., Krause, F., Dilba, P., Tacke, M., and Karl, J. (2010) A combination of serum markers for the early detection of colorectal cancer. *Clin. Cancer Res.* **16**, 6111–6121
82. Polanski, M., and Anderson, N. L. (2006) A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights* **16**, 1–48
83. Locker, G. Y., Hamilton, S., Harris, J., Jessup, J. M., Kemeny, N., Macdonald, J. S., Somerfield, M. R., Hayes, D. F., and Bast, R. C. Jr (2006) ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–5327
84. Wu, X. Z., Ma, F., and Wang, X. L., (2010) Serological diagnostic factors for liver metastasis in patients with colorectal cancer. *World J. Gastroenterol.* **16**, 4084–4088
85. Oue, N., Kuniyasu, H., Noguchi, T., Sentani, K., Ito, M., Tanaka, S., Setoyama, T., Sakakura, C., Natsugoe, S., and Yasui, W. (2007) Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis. *Oncology* **72**, 371–380

86. Marti, J., Fuster, J., Solà, A. M., Hotter, G., Molina, R., Pelegrina, A., Ferrer, J., Deulofeu, R., Fondevila, C., and García-Valdecasas, J. C. (2013) Prognostic value of serum neutrophil gelatinase-associated lipocalin in metastatic and nonmetastatic colorectal cancer. *World J. Surg.* **37**, 1103–1109
87. Bünger, S., Haug, U., Kelly, M., Posorski, N., Klempt-Giessing, K., Cartwright, A., Fitzgerald, S. P., Toner, V., McAleer, D., Gemoll, T., Laubert, T., Büning, J., Fellermann, K., Bruch, H. P., Roblick, U. J., Brenner, H., von Eggeling, F., and Habermann, J. K. (2012) A novel multiplex-protein array for serum diagnostics of colon cancer: a case-control study. *BMC Cancer* **12**, 393
88. Yu, J., Li, X., Zhong, C., Li, D., Zhai, X., Hu, W., Guo, C., Yuan, Y., and Zheng, S. (2016) High-throughput proteomics integrated with gene microarray for discovery of colorectal cancer potential biomarkers. *Oncotarget* **7**, No. 46
89. de Boer, N. K., de Meij, T. G., Oort, F. A., Ben Larbi, I., Mulder, C. J., van Bodegraven, A. A., and van der Schee, M. P. (2014) The scent of colorectal cancer: detection by volatile organic compound analysis. *Clin. Gastroenterol. Hepatol.* **12**, 1085–1089
90. Sonoda, H., Kohnoe, S., Yamazato, T., Satoh, Y., Morizono, G., Shikata, K., Morita, M., Watanabe, A., Morita, M., Kakeji, Y., Inoue, F., and Maehara, Y. (2011) Colorectal cancer screening with odour material by canine scent detection. *Gut* **60**, 814–819
91. Hirotsu, T., Sonoda, H., Uozumi, T., Shinden, Y., Mimori, K., Maehara, Y., Ueda, N., and Hamakawa, M. (2015) A highly accurate inclusive cancer screening test using *Caenorhabditis elegans* scent detection. *PLoS ONE* **10**, e0118699
92. Stenvang, J., Kümler, I., Nygård, S. B., Smith, D. H., Nielsen, D., Brünner, N., and Moreira, J. M. A. (2013) Biomarker-guided repurposing of chemotherapeutic drugs for cancer therapy: a novel strategy in drug development. *Front. Oncol.* **3**, Article 313
93. Meisenberg, C., Gilbert, D. C., Chalmers, A., Haley, V., Gollins, S., Ward, S. E., and El-Khamisy, S. F. (2015) Clinical and cellular roles for TDP1 and TOP1 in modulating colorectal cancer response to irinotecan. *Mol. Cancer Ther.* **14**, 575–585
94. Kolquist, K. A., Ellisen, L. W., Counter, C. M., Meyerson, M., Tan, L. K., Weinberg, R. A., Haber, D. A., and Gerald, W. L. (1998) Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. *Nat. Genet.* **19**, 182–186
95. Mailliard, W. S., Haigler, H. T., and Schlaepfer, D. D. (1996) Calcium-dependent binding of S100C to the N-terminal domain of annexin I. *J. Biol. Chem.* **271**, 719–725
96. Deloulme, J. C., Assard, N., Mbele, G. O., Mangin, C., Kuwano, R., and Baudier, J. (2000) S100A6 and S100A11 are specific targets of the calcium- and zinc-binding S100B protein in vivo. *J. Biol. Chem.* **275**, 35302–35310
97. Sakaguchi, M., Miyazaki, M., Takaishi, M., Sakaguchi, Y., Makino, E., Kataoka, N., Yamada, H., Namba, M., and Huh, N. H. (2003) S100C/A11 is a key mediator of Ca²⁺-induced growth inhibition of human epidermal keratinocytes. *J. Cell Biol.* **163**, 825–835
98. Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar V. U., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R., Paccanaro, A., Marcotte, E. M., and Emili, A. (2012) A census of human soluble protein complexes. *Cell* **150**, 1068–1081
99. Baek, J. Y., Yeo, H. Y., Chang, H. J., Kim, K. H., Kim, S. Y., Park, J. W., Park, S. C., Choi, H. S., Kim, D. Y., and Oh, J. H. (2013) Serpin B5 is a CEA-interacting biomarker for colorectal cancer. *Int. J. Cancer* **134**, 1595–1604